The Third Generation ACD Model: A Semiparametric Approach

By

Pipat Wongsaart[1], Jiti Gao[2] and David E. Allen[3]

[1]School of Mathematics and Statistics, University of Western Australia
[2]School of Economics, University of Adelaide
[3]School of Accounting, Finance and Economics, Edith Cowan University

School of Accounting, Finance and Economics & FEMARC Working Paper Series
Edith Cowan University
June 2008
Working Paper 0806

Correspondence author:

David E. Allen
School of Accounting, Finance and Economics
Faculty of Business and Law
Edith Cowan University
Joondalup, WA 6027
Australia
Phone: +618 6304 5471
Fax: +618 6304 5271
Email: d.allen@ecu.edu.au

ABSTRACT

Having concluded that thus far the question about the most appropriate type of nonlinear ACD model has not been satisfactorily answered, we intend to develop a novel ACD modelling methodology based on an iterative estimation algorithm and a semiparametric autoregressive process that not only allows the data to speak for itself, but also is robust across datasets without relying on some computational factors, such as the hypothesis about the probability density function of the standardised durations. We propose in this paper the Semiparametric ACD (SP-ACD) model, which can be considered a starting point of such a development. To address the problem about the unobservability of the conditional durations in practice, the current paper devises an iterative algorithm to estimate the unknown conditional duration process. In such a circumstance, it is essential to provide not only the mathematical justification of the estimation scheme, but also sound asymptotic results about the semiparametric and the adaptive data-driven estimators. This paper focuses mainly on the former and also on a number of simulation experiments, while the later is left for future study.

Keywords: Semiparametric ACD model; Conditional durations; Simulations

1. Introduction

Since the seminal work of Engle and Russell (1998), the modelling of financial data at transaction level has become an ongoing topic in the area of financial econometrics. The ultimate high frequency data in finance are transaction-by-transaction or trade-by-trade data in security markets where time is often measured in seconds. Transaction data possesses a number of unique characteristics that do not appear in lower frequencies.[1] The most salient feature, however, is that they are fundamentally irregularly spaced. This feature challenges researchers as standard econometric techniques refined over the years are no longer applicable. Moreover, recent models from the market microstructure literature[2] argue that time may convey information and, therefore, should also be modelled.

Motivated by these considerations, Engle and Russell (1998) developed the Autoregressive Conditional Duration (ACD) model such that the arrival times are modelled as random variables, which follow a point process. Apparently, the ACD model and the GARCH model of Bollerslev (1986) share several common features. Both models rely on a similar economic motivation, i.e. the clustering of news and financial events in the markets. Due to their structural similarity, a number of researchers have subsequently proposed numerous parametric extensions to the ACD model in a similar manner to those of the GARCH, for example, the logarithmic ACD (Log-ACD) model of Bauwens and Giot (2000). There are also some other parametric extensions of the ACD model that abound in the literature. Fernandes and Grammig (2006), for example, consider a family of ACD-type models that relies on asymmetric responses to shocks and on a Box-Cox transformation to the conditional duration process. Their family encompasses most parametric ACD models in the literature, though there are a few exceptions. Zhang, Russell, and Tsay (2001) argue for a nonlinear version based on self-exciting threshold ACD processes, whereas Meitz and Terässvirta (2006) propose the smooth transition and the time-varying ACD models.[3]

[1] See Tsay (2005) for an excellent review.
[2] See, for example, Hasbrouck (1991), and Easley and O'Hara, M. (1992).
[3] See Pacurar (2008) for an excellent survey on both the theoretical and empirical works that have been done on the ACD model.

The important motivation for developing these models was to allow for some additional flexibility mainly in order to address the issue of nonlinearity that was raised in a number of studies.[4] However, thus far the question about the most appropriate type of nonlinear ACD model has not been efficiently addressed. In our view, the scope of nonlinearity offered by the above parametric extensions is still too limited for some, while excessively sophisticated in the others. We believe that not only a better model should be able to fully capture nonlinear influences of the past information on the current conditional duration, but it also should be flexible in the sense that data is allowed to speak for itself and that it is robust across duration processes. Furthermore, the performance of the model should not depend on some computational factors, such as the hypothesis about the probability density function of the standardised durations.

Motivated by these considerations, we put forward a novel methodology of ACD modelling, which can be considered the third generation approach, based on iterative estimation algorithms and semiparametric autoregressive processes. The current paper can be considered a starting point in which we discuss the Semiparametric ACD (SP-ACD) model. Clearly, semiparametrics should bring about much needed flexibility. Fernandes, Medeiros and Veiga (2008), for example, proposed what called the Functional Coefficient ACD (FC-ACD), which, by increasing the number of regimes to infinity, can acts as a universal neural-network approximation. The main difficulty of estimating an ACD model using semiparametric techniques directly resides in the unobservability of the conditional durations in practice. To address this problem, the current paper devises an iterative algorithm to estimate the unknown conditional duration process. In such a circumstance, it is essential to provide not only the mathematical justification of the estimation scheme, but also sound asymptotic results about the semiparametric and the adaptive data-driven estimators. The current paper will focus mainly on the former and also on a number of simulation experiments, while the later will be left for future study.

[4] See, for example, Engle and Russell (1998), Dufour and Engle (2000), Zhang, Russell and Tsay (2001), and Fernandes and Grammig (2006).

The following section develops the statistical underpinning for the SP-ACD model and Section 3 presents the basic construction of the above-mentioned estimation algorithm. Section 4 discusses asymptotic properties of the SP-ACD model and Section 5 considers a number of illustrative examples and gives simulation results. The paper concludes with a discussion, while proofs are given in the appendix.

2 The SP-ACD Model

Let consider a stochastic process that is simply a sequence of times $\{t_0, t_1, \ldots, t_n, \ldots\}$ with $t_0 < t_1 < \cdots < t_n \ldots$. In this case, $x_i = t_i - t_{i-1}$ defines the intervals between two arrival times, which is commonly known as the durations. In this paper, we consider the duration process $\{x_i, i \in \mathbb{Z}\}$, which is assumed to be a nonnegative and strictly stationary stochastic process adapted to the filtration $\{\Omega_i, i \in \mathbb{Z}\}$ with $\Omega_i = \sigma(\{x_s; s \le i\})$, of the form

$$x_i = \psi_i \varepsilon_i \tag{2.1}$$

where

$$E\left[x_i \middle| x_{i-1}, \ldots, x_1, \psi_{i-1}, \ldots, \psi_1\right] = f\left(x_{i-1}, \ldots, x_1, \psi_{i-1}, \ldots, \psi_1\right) \equiv \psi_i \tag{2.2}$$

is the conditional expectation of the ith duration, which is dependent upon the past information, and $\{\varepsilon_i; i \in \mathbb{Z}\}$ is an i.i.d. innovation series, independent of $\{x_s; s < i\}$, with unit mean and variance, and a finite fourth moment. Statistically ε can follow any distribution function, $F(\varepsilon)$, such that $P(\varepsilon < 0) = 0$. One of the most popular choices among ACD studies is the generalised Gamma distribution with a scale parameter α, and the shape parameters (β, κ) all greater than zero, while nested within it are the Gamma distribution (where $\alpha = 1$) and the Weibull distribution (where $\kappa = 1$).[5]

The basic ACD model as proposed by Engle and Russell (1998) relies on a linear parameterisation of (2.2) in which ψ_i depends on p past durations and q past expected durations as

$$\psi_i = \omega + \sum_{j=1}^{p} \gamma_j x_{i-j} + \sum_{k=1}^{q} \beta_k \psi_{i-k} . \tag{2.3}$$

[5] See Grammig and Maurer (2000), Allen et al. (2006) and Vuorenmaa (2006) for discussion on these and other distribution functions.

In literature, this is often referred to as the $\text{ACD}(p,q)$ model by which sufficient conditions to ensure the positivity of ψ_i are, $\omega > 0$, $\gamma_i > 0\ \forall i = 1,\ldots,p$ and $\beta_i \geq 0\ \forall i = 1,\ldots,q$. Note that these conditions are identical to that of the GARCH model, which ensure that the conditional variance is positive.

Despite the evidence of nonlinearity reported in a various studies,[6] the question about the type of nonlinear ACD model that is the most appropriate, has not yet been satisfactorily addressed. As an alternative, the current paper proposes the $\text{SP-ACD}(\text{p,q})$ model of the form

$$\begin{aligned} E\left[x_i \left| x_{i-1},\ldots,x_{i-p},\psi_{i-1},\ldots,\psi_{i-q}\right.\right] &= f\left(x_{i-1},\ldots,x_{i-p},\psi_{i-1},\ldots,\psi_{i-q}\right) \\ &= \sum_{j=1}^{p}\gamma_j x_{i-j} + \sum_{k=1}^{q} g_k\left(\psi_{i-k}\right) \\ &\equiv \psi_i . \end{aligned} \tag{2.4}$$

However, in order to present the main ideas and methodology without unnecessary complication in our discussion, here attention will be restricted only to a special case of (2.4), namely

$$E\left[x_i \left| x_{i-1},\psi_{i-1}\right.\right] = f\left(x_{i-1},\psi_{i-1}\right) = \gamma x_{i-1} + g\left(\psi_{i-1}\right) \equiv \psi_i \tag{2.5}$$

where γ is an unknown parameter and $g(\cdot)$ is an unknown function on the real line. The strict stationarity assumption being imposed on $\{x_i, i \in \mathbb{Z}\}$ suggests that γ and $g(\cdot)$ must satisfy a number of conditions, including (1.5) and those of Theorem 3.1 of An and Huang (1996).

To derive the estimators of γ and g, we first write (2.1) as

$$x_i = f\left(x_{i-1},\psi_{i-1}\right) + \eta_i , \tag{2.6}$$

where $\eta_i = f\left(x_{i-1},\psi_{i-1}\right)\left(\varepsilon_i - 1\right)$ is a martingale difference series. Furthermore, using the functional form given in (2.5), we have

$$x_i = \gamma x_{i-1} + g\left(\psi_{i-1}\right) + \eta_i \tag{2.7}$$

where, in this case,

$$g\left(\psi_{i-1}\right) = E\left[\left(x_i - \gamma x_{i-1}\right)\left|\psi_{i-1}\right.\right] = E\left(x_i\left|\psi_{i-1}\right.\right) - \gamma E\left(x_{i-1}\left|\psi_{i-1}\right.\right) = g_1\left(\psi_{i-1}\right) - \gamma g_2\left(\psi_{i-1}\right). \tag{2.8}$$

[6] See, for example, Dufour and Engle (2000), Zhang, Russell and Tsay (2001), and Fernandes and Grammig (2006).

In addition, a slight simplification of (2.7) leads to

$$x_i = \gamma z_i + g_1(\psi_{i-1}) + \eta_i \tag{2.9}$$

where $z_i = x_{i-1} - g_2(\psi_{i-1})$. If γ is known to be the true parameter, then the natural estimates of g_j and g for a given γ are

$$\hat{g}_{1,h}(\psi_{i-1}) = \sum_{s=2}^{T} W_{s,h}(\psi_{i-1}) x_s \,, \tag{2.10}$$

$$\hat{g}_{2,h}(\psi_{i-1}) = \sum_{s=2}^{T} W_{s,h}(\psi_{i-1}) x_{s-1} \,, \tag{2.11}$$

and

$$\hat{g}_h(\psi_{i-1}) = \hat{g}_{1,h}(\psi_{i-1}) - \gamma \hat{g}_{2,h}(\psi_{i-1}) \tag{2.12}$$

where $W_{s,h}(\psi_{i-1})$ is a probability weight function depending on $\psi_1, \psi_2, \ldots, \psi_{T-1}$ and the number T of observations. Hereafter, let $\overline{x}_i = x_i - \hat{g}_{1,h}(\psi_{i-1})$, then the kernel-weighted least squares estimator of γ can be found by minimising

$$\sum_{i=2}^{T} \left\{ \overline{x}_i - \gamma u_i \right\}^2 \tag{2.13}$$

to obtain

$$\hat{\gamma}_h - \gamma = \left[\sum_{i=2}^{T} u_i^2 \right]^{-1} \left\{ \sum_{i=2}^{T} u_i \cdot \eta_i + \sum_{i=2}^{T} u_i \cdot \overline{g}_h(\psi_{i-1}) \right\} \tag{2.14}$$

where $u_i = x_{i-1} - \hat{g}_{2,h}(\psi_{i-1})$ and $\overline{g}_h(\psi_{i-1}) = g(\psi_{i-1}) - \hat{g}_h(\psi_{i-1})$.

Furthermore, if $\sigma^2 = E\left[\eta_1^2\right]$ is unknown, it can be estimated by

$$\hat{\sigma}^2(h) = \frac{1}{(T-1)} \sum_{i=2}^{T} \left\{ x_i - \hat{\gamma}_h x_{i-1} - \hat{g}_h^*(\psi_{i-1}) \right\}^2 \tag{2.15}$$

where $\hat{g}_h^*(\psi_{i-1}) = \hat{g}_{1,h}(\psi_{i-1}) - \hat{\gamma}_h \hat{g}_{2,h}(\psi_{i-1})$.

The quality of the proposed estimators can be measured by the average squared error (ASE). However, in our analysis, computation of the ASE must also takes into account the fact that the true conditional durations must themselves be estimated. Therefore, this paper defines the ASE as follows

$$\mathfrak{D}(h) = \frac{1}{(T-1)} \sum_{i=2}^{T} \left[\left\{ \hat{\gamma}_h x_{i-1} + \hat{g}_{1,h}(\hat{\psi}_{i-1}) - \hat{\gamma}_h \hat{g}_{2,h}(\hat{\psi}_{i-1}) \right\} - \left\{ \gamma x_{i-1} + g_1(\psi_{i-1}) - \gamma g_2(\psi_{i-1}) \right\} \right]^2 \omega(\hat{\psi}_{i-1}, \psi_{i-1})$$

where $\hat{\psi}$ is the estimate of the conditional duration computed using the estimation scheme, which is explained in detail below, and $\omega(\hat{\psi}_i, \psi) = w(\hat{\psi}_i) w(\psi_i)$ such that w is a weight function.

Note that this paper considers only the case where $W_{s,h}$ is a kernel weight function

$$W_{s,h}(y) = K_h(y - \psi_{s-1}) \Big/ \sum_{i=2}^{T} K_h(y - \psi_{i-1})$$

where $K_h(\cdot) = h^{-1} K(\cdot/h)$, K is a real-valued kernel function satisfying Assumptions 4a below and $h = h_T \in H_T = \left[a_1 T^{-1/5-c_1}, b_1 T^{-1/5+c_1} \right]$ in which $0 < a_1 < b_1 < \infty$ and $0 < c_1 < 1/20$.

An obvious advantage of the above SP-ACD model over its ACD counterpart is the additional flexibility by which the linear specification is nested as a special case. If indeed the current duration does depend on the past conditional durations with some unknown nonlinear relationship, the estimate of the parameter γ in (2.14) would be much more accurate than those offered by the parametric extensions.

Another important benefit is the fact that it also makes possible a choice of modelling

$$x_i = \phi(x_{i-1}) + \lambda \psi_{i-1} + e_i \tag{2.16}$$

where ϕ is an unknown function on the real line, λ is an unknown parameter and e_i is a sequence of i.i.d. random errors with mean zero and finite variance. Statistical tests conducted by Engle and Russell (1998) on IBM trade duration data have found significant evidence of a nonlinear relationship between conditional durations and past information set. In particular, the authors report that on average the expected durations from the linear model are too large after the very short as well as the very long durations. It is obvious that the model in (2.16) does enable a nonlinear relationship such that very short durations would have a bigger impact on decreasing expectations, while long durations would have a reduced impact on lengthening durations. Therefore, the SP-ACD of this kind could potentially be very useful empirically.

The following section discusses the iterative scheme adapted in this paper in order to compute an estimate of the true, but unobservable, conditional duration.

3. The Computational Algorithm

In this section, we first present the basic construction of the scheme, then discuss its theoretical justification. Note that mathematical proofs of the results in this section are relegated to the appendix.

Let us assume that we have a set of data samples $\{x_i; 1 \le i \le T\}$, ideally from the generating process described by (2.7). The estimation algorithm is constructed to include five important steps as follows.

<u>Step 1</u>: Choose the starting values for the vector of the T conditional means. Index this values by a zero, i.e. $\{\psi_{i,0}\}$ and set $m = 1$.

<u>Step 2</u>: Compute $\hat{f}_{h,m}$, which includes $\hat{\gamma}_{h,m}$ and $\hat{g}_{h,m}$, based on $\{x_{i-1}; 2 \le i \le T\}$ and the estimates of the conditional durations as computed in the previous step, i.e. $\{\hat{\psi}_{i-1,m-1}; 2 \le i \le T\}$, using semiparametric regression technique.

<u>Step 3</u>: Compute $\left\{\hat{\psi}_{i,m} = \hat{f}_{h,m}\left(x_{i-1}, \hat{\psi}_{i-1,m-1}; 2 \le i \le T\right)\right\}$ and also select some sensible value for $\hat{\psi}_{1,m}$, which cannot be computed recursively.

<u>Step 4</u>: If $m < M$, where M is a pre-specified maximum number of iterations, then increment m and return to Step 2.

<u>Step 5:</u> When $m = M$, perform the final semiparametric regression of x_{i-1} and $\hat{\psi}_{i,m}$ to obtain the final estimates $\hat{\psi}_i = \hat{f}_h\left(x_{i-1}, \hat{\psi}_{i-1,m}\right)$. Often, the performance of the algorithm can be improved by averaging over the final K of M iterations to obtain

$$\hat{\psi}_{i,A} = \left(\frac{1}{K}\right) \sum_{m=M-K+1}^{M} \hat{\psi}_{i,m}, \tag{3.1}$$

then perform the final semiparametric regression of x_{i-1} and $\hat{\psi}_{i-1,A}$, which leads ultimately to the final estimates $\hat{\psi}_i = \hat{f}\left(x_{i-1}, \hat{\psi}_{i-1,A}\right)$.

We will now discuss the theoretical justification of the above iterative estimation scheme. Let us first introduce the following terms and notations. The $\text{SPL-ACD}(1,1)$ model in (2.5) suggests that the estimates of conditional durations at the mth $(m>1)$ iteration can be defined as

$$\hat{\psi}_{i,n,m} = \hat{\gamma}_h x_{i-1} + \hat{g}_{1,h}\left(\hat{\psi}_{i-1,n,m-1}\right) - \hat{\gamma}_h \hat{g}_{2,h}\left(\hat{\psi}_{i-1,n,m-1}\right) \tag{3.2}$$

where

$$\hat{g}_{j,h}\left(\cdot\right) = \sum_{s=2}^{T} W_{s,h}\left(\cdot\right) x_{s-j+1} \quad \left(j = 1,2\right) \tag{3.3}$$

are the nonparametric kernel estimates of g_j with $W_{s,h}$ being a probability weight function depending on $\hat{\psi}_{1,n,m-1}, \hat{\psi}_{2,n,m-1}, \ldots, \hat{\psi}_{T-1,n,m-1}$, and $\hat{\gamma}_h$ is a kernel-weighted estimator for the unknown parameter γ. Furthermore,

$$E\left[x_i \middle| x = x_{i-1}, \psi = \hat{\psi}_{i-1,n,m-1}\right] = \gamma x_{i-1} + g_1\left(\hat{\psi}_{i-1,n,m-1}\right) - \gamma g_2\left(\hat{\psi}_{i-1,n,m-1}\right) \equiv \overline{\psi}_{i,n,m} \tag{3.4}$$

$$E\left[x_i \middle| x = x_{i-1}, \psi = \psi_{i-1,n,m-1}\right] = \gamma x_{i-1} + g_1\left(\psi_{i-1,n,m-1}\right) - \gamma g_2\left(\psi_{i-1,n,m-1}\right) \equiv \psi_{i,n,m} \tag{3.5}$$

$$g_1\left(\hat{\psi}_{i-1,n,m-1}\right) = E\left[x_i \middle| \hat{\psi}_{i-1,n,m-1}\right],\; g_2\left(\hat{\psi}_{i-1,n,m-1}\right) = E\left[x_{i-1} \middle| \hat{\psi}_{i-1,n,m-1}\right], \tag{3.6}$$

and

$$g_1\left(\psi_{i-1,n,m-1}\right) = E\left[x_i \middle| \psi_{i-1,n,m-1}\right],\; g_2\left(\psi_{i-1,n,m-1}\right) = E\left[x_{i-1} \middle| \psi_{i-1,n,m-1}\right] \tag{3.7}$$

where $i = 1,\ldots,T$, $n = 1,\ldots,N$ and $m = 1,\ldots,M$. While, $\overline{\psi}_{i,n,m}$ in (3.4) represents the true conditional expectation as a function of x_{i-1} and the estimate $\hat{\psi}_{i-1,n,m-1}$, $\psi_{i,n,m}$ in (3.5) are the population quantities corresponding to the estimates $\hat{\psi}_{i,n,m}$ of the algorithm where $\psi_{i,n,0}$ are some starting values assumed to be elements of Ω_{i-1} for all $i \in \mathbb{Z}$, i.e. they are independent from $\left\{\varepsilon_s; s \le i\right\}$. Hereafter, let us denote by $\|\cdot\|_2$, the $L_{2\text{-norm}}$ such that

$$\|\mathrm{Y}\|_2 = \left(E|Y|^2\right)^{1/2}. \tag{3.8}$$

With terms as defined in (3.2) to (3.7), we now state and bound the quantity of interest as

$$\left\|\hat{\psi}_{i,n,m} - \psi_i\right\|_2 \le \left\|\hat{\psi}_{i,n,m} - \overline{\psi}_{i,n,m}\right\|_2 + \left\|\overline{\psi}_{i,n,m} - \psi_{i,n,m}\right\|_2 + \left\|\psi_{i,n,m} - \psi_i\right\|_2, \tag{3.9}$$

where

$$\left\|\hat{\psi}_{i,n,m}-\overline{\psi}_{i,n,m}\right\|_2=\Big\|\left(\hat{\gamma}_h x_{i-1}+\hat{g}_{1,h}\left(\hat{\psi}_{i-1,n,m-1}\right)-\hat{\gamma}_h\hat{g}_{2,h}\left(\hat{\psi}_{i-1,n,m-1}\right)\right)$$
$$-\left(\gamma x_{i-1}+g_1\left(\hat{\psi}_{i-1,n,m-1}\right)-\gamma g_2\left(\hat{\psi}_{i-1,n,m-1}\right)\right)\Big\|_2, \tag{3.10}$$

$$\left\|\overline{\psi}_{i,n,m}-\psi_{i,n,m}\right\|_2=\left\|\left(g_1\left(\hat{\psi}_{i-1,n,m-1}\right)-g_1\left(\psi_{i-1,n,m-1}\right)\right)-\gamma\left(g_2\left(\hat{\psi}_{i-1,n,m-1}\right)-g_2\left(\psi_{i-1,n,m-1}\right)\right)\right\|_2, \tag{3.11}$$

and

$$\left\|\psi_{i,n,m}-\psi_i\right\|_2=\left\|\left(g_1\left(\psi_{i-1,n,m-1}\right)-g_1\left(\psi_{i-1}\right)\right)-\gamma\left(g_2\left(\psi_{i-1,n,m-1}\right)-g_2\left(\psi_{i-1}\right)\right)\right\|_2. \tag{3.12}$$

It is obvious that, while the first error term on the right side of (3.9) quantifies the estimation errors of a one-step semiparametric regression at the mth iteration, the third term represents the case without estimation error such that $\psi_{i,n,0}$ are some starting values. This suggests, therefore, that in this case $\left\|\psi_{i,0}-\psi_i\right\|_2$ quantifies the error due to wrong starting values. Furthermore, by denoting

$$\hat{\psi}_{i,\kappa}=\left(\hat{\psi}_{i,1,\kappa},\ldots,\hat{\psi}_{i,N,\kappa}\right)^\tau,\ \overline{\psi}_{i,\kappa}=\left(\overline{\psi}_{i,1,\kappa},\ldots,\overline{\psi}_{i,N,\kappa}\right)^\tau \text{ and } \psi_{i,\kappa}=\left(\psi_{i,1,\kappa},\ldots,\psi_{i,N,\kappa}\right)^\tau \tag{3.13}$$

for $1\le\kappa\le m-1$, hereafter, let us define

$$\Delta_{1\kappa}^{(L_2)}=:\sup_{\kappa\ge1}\left\|\hat{\psi}_{i,\kappa}-\overline{\psi}_{i,\kappa}\right\|_2$$

and (3.14)

$$\Delta_{2\kappa}^{(L_2)}=:\sup_{\kappa\ge1}\left\|\overline{\psi}_{i,\kappa}-\psi_{i,\kappa}\right\|_2.$$

We will now present assumptions, which are crucial to the statistical justification of the above estimation algorithm, and the main results of this section.

Assumption 3a:

Functions $g_j(\psi)$ $(j=1,2)$ on the real line satisfy the following Lipschitz type condition:

$$\left|g_j(\psi+v)-g_j(\psi)\right|\le\varphi_j(\psi)\cdot|v| \tag{3.15}$$

for each given ψ and all $v\in S$ (any compact subset of $\mathbb{R}^1$) by which $\varphi_j(\psi)$ are measurable functions such that

$$\left(\left\|\varphi_1(\psi)\right\|_2\right)^{1/2}+\left(\left\|\varphi_2(\psi)\right\|_2\right)^{1/2}<G \tag{3.16}$$

for some $0<G<1$.

Assumption 3b:

$$V_1 := \|\psi_i\|_2 < \infty \text{ and } V_2 := \|\psi_{i,0}\|_2 < \infty . \tag{3.17}$$

Theorem 3a:

Assume that Assumptions 3a and 3b hold. Then

$$\Delta_{1m}^{(L_2)} = \left(\frac{C_1}{Th} + C_2 h^4 + o_P \left\{ \frac{1}{Th} + h^4 \right\} \right)^{1/2} \text{ uniformly over } h \in H_T \tag{3.18}$$

where $c_i \left(i = 1,2 \right)$ *are some positive constants and* $\Delta_{1m}^{(L_2)}$ *is as defined in (3a.3) below.*

Theorem 3b:

Assume that $\left\{ x_i ; i \in \mathbb{Z} \right\}$ *is as in (2.7). Then for all* $i \geq 1$

$$\left\| \psi_{i,n,m} - \psi_i \right\|_2 \leq G^m \left(V_1 + V_2 \right) \tag{3.19}$$

for some $0 < G < 1$.

Theorem 3c:

Assume that $\left\{ x_i ; i \in \mathbb{Z} \right\}$ *is as in (2.7). Then for all* $i \geq 1$

$$\left\| \overline{\psi}_{i,n,m} - \psi_{i,n,m} \right\|_2 \leq \sum_{j=1}^{m-1} G^j \Delta_{1\kappa}^{(L_2)} \left\{ 1 + 2 \left(\Delta_{2\kappa}^{(L_2)} / \Delta_{1\kappa}^{(L_2)} \right)^{1/2} \right\} \tag{3.20}$$

for some $0 < G < 1$.

Theorem 3d:

Assume that $\left\{ x_i ; i \in \mathbb{Z} \right\}$ *is as in (2.7) and that Assumptions 3a, 3b and 4a hold. Then*

$$\left\| \hat{\psi}_{i,n,m} - \psi_i \right\|_2^2 = O\left(\Delta_{1m}^{(L_2)} \right)^2 = O\left(\frac{c_1}{Th} + c_2 h^4 + o_P \left\{ \frac{1}{Th} + h^4 \right\} \right) \tag{3.21}$$

uniformly over $h \in H_T$.

4. Asymptotic Theory

Asymptotic theory of the above semiparametric estimation has been considered in a number of studies, for example Engle et. al. (1986), Heckman (1986), Härdle et al. (2000), Gao and Yee (2000), and Gao (2006). Nonetheless, the fact that here the conditional durations themselves must also be estimated significantly differentiate the current study from the previous ones. Therefore, it is our intention to ensure that in this case the semiparametric and the adaptive data-driven estimations of models (2.7) and (2.16) are of sound asymptotic characters. In this introductory paper, however, we only state expected asymptotic results without proofs, while focusing more on Section 5 where we will illustrate how well the above estimation procedure works numerically and practically.

Let us begin our discussion in this section by rewriting the kernel-weighted LS estimators $\hat{\gamma}_h$ of γ in (2.14) using $\hat{\psi}_i$, which hereafter represents our estimates of the conditional durations, as follows

$$\hat{\gamma}_h - \gamma = \left(\sum_{i=2}^{T} \hat{z}_i^2 \right)^{-1} \left(\sum_{i=2}^{T} \hat{z}_i \cdot \eta_i + \sum_{i=2}^{T} \hat{z}_i \cdot \hat{\bar{g}}_h \right) \tag{4.1}$$

where $\hat{z}_i = x_{i-1} - \hat{g}_{2,h}\left(\hat{\psi}_{i-1}\right)$ and $\hat{\bar{g}}_h = g\left(\psi_{i-1}\right) - \hat{g}_h\left(\hat{\psi}_{i-1}\right)$. We list below assumptions, which are considered quite common among this type of study.

Assumption 4a:

(i) Assume that the processes $\left(\psi_i : i \geq 1\right)$ are strictly stationary and α-mixing with mixing coefficient $\alpha(T) = Cq^T$, for some $0 < C < \infty$ and $0 < q < 1$, and that $\{\psi_i\}$ has a common marginal density $f(\cdot)$ where $f(\cdot)$ has a compact support containing S with two continuous derivatives on the interior of S.

(ii) Assume that $\{\eta_i\}$ is a sequence of independently and identically distributed random processes with $E(\eta_i) = 0$ and $E\left(\eta_i^2\right) = \sigma_0^2 < \infty$, and that η_s are independent of x_n and ψ_n for all $s \geq n$.

(iii) Assume that the weight function ω is bounded and that its support S is compact.

(iv) Assume that the functions $g_j(\cdot)$, for $j = 1, 2$, have two continuous derivatives on the interior of S.

(v) Assume that the kernel function, K, is symmetric, Lipschitz continuous and has an absolutely integrable Fourier transform and that it is a bounded probability function with $\int K(u)\,du = 1$, $K(\cdot) \geq 0$ and $\int_{-\infty}^{\infty} u^2 K(u)\,du < \infty$.

(vi) For any integer $k \geq 1$, $E\left|\psi_i\right|^k < \infty$.

Note that Assumption 4a(i) is similar to (C.1) and (C.8) of Härdle and Vieu (1992). The remaining assumptions are statistically justified due to the stationarity of ψ_i. Finally, within the context of a one-step semiparametric regression, Gao and Yee (2000), and Härdle and Vieu (1992) show that Assumption 4a(vi) is required in order to prove that the data-driven bandwidth is asymptotically optimal.

Proposition 4a:

Assume that Assumptions 4a holds. Then the following holds uniformly over $h \in H_T$:

$$\sqrt{T}\left\{\hat{\gamma}_h - \gamma\right\} \to N\left[0, \sigma^2/\sigma_2^2\right] \tag{4.2}$$

where $\sigma_2^2 = E\left\{x_{i-1} - E\left(x_{i-1}\middle|\psi_{i-1}\right)\right\}^2 > 0$.

In view of (4.1), in order to prove Proposition 4a, we must show that

$$T^{-1/2}\sum_{i=2}^{T}\hat{z}_i \cdot \eta_i \to N\left[0, \sigma^2/\sigma_2^2\right], \tag{4.3}$$

$$\sum_{i=2}^{T}\left\{g_2\left(\psi_{i-1}\right) - \hat{g}_{2,h}\left(\hat{\psi}_{i-1}\right)\right\}\eta_i = o_P\left(T^{1/2}\right), \tag{4.4}$$

$$\sum_{i=2}^{T}\left\{g_h\left(\psi_{i-1}\right) - \hat{g}_h\left(\hat{\psi}_{i-1}\right)\right\}\hat{z}_i = o_P\left(T^{1/2}\right), \tag{4.5}$$

and

$$\sum_{i=2}^{T}\left\{g_2\left(\psi_{i-1}\right) - \hat{g}_{2,h}\left(\hat{\psi}_{i-1}\right)\right\}\left\{g_h\left(\psi_{i-1}\right) - \hat{g}_h\left(\hat{\psi}_{i-1}\right)\right\} = o_P\left(T^{1/2}\right) \textit{ uniformly over } h \in H_T \tag{4.6}$$

where $g_h\left(\psi_{i-1}\right) - \hat{g}_h\left(\hat{\psi}_{i-1}\right) = \left\{g_1\left(\psi_{i-1}\right) - \hat{g}_{1,h}\left(\hat{\psi}_{i-1}\right)\right\} - \gamma\left\{g_2\left(\psi_{i-1}\right) - \hat{g}_{2,h}\left(\hat{\psi}_{i-1}\right)\right\}$.

It is not difficult to show, in an usual semiparametric case, that

$$\lim_{T\to\infty}\frac{1}{T}\sup_{h\in H_T}\left\{\sum_{i=2}^{T}u_i^2\right\} = \sigma_2^2, \tag{4.7}$$

therefore, for the leading term in (4.3) to be true we only have to show that

$$\frac{1}{T}\sup_{h\in H_T}\left(\sum_{i=2}^{T}\hat{z}_i^2\right) \xrightarrow{P} \frac{1}{T}\sum_{i=2}^{T}u_i^2 . \tag{4.8}$$

Also, the remaining terms can be proved using the results of Theorem 3d and the fact that

$$\frac{1}{T}\sum_{i=2}^{T}\left\{\hat{g}_{j,h}\left(\hat{\psi}_{i-1}\right)-g_j\left(\psi_{i-1}\right)\right\}^2\omega\left(\psi_{i-1},\hat{\psi}_{i-1}\right)$$
$$=\frac{1}{T}\sum_{i=2}^{T}\left\{\hat{g}_{j,h}\left(\hat{\psi}_{i-1}\right)-\hat{g}_{j,h}\left(\psi_{i-1}\right)\right\}^2\omega\left(\psi_{i-1},\hat{\psi}_{i-1}\right)+\frac{1}{T}\sum_{i=2}^{T}\left\{\hat{g}_{j,h}\left(\psi_{i-1}\right)-g_j\left(\psi_{i-1}\right)\right\}^2\omega\left(\psi_{i-1},\hat{\psi}_{i-1}\right) \tag{4.9}$$
$$+\frac{2}{T}\sum_{i=2}^{T}\left\{\hat{g}_{j,h}\left(\hat{\psi}_n\right)-\hat{g}_{j,h}\left(\psi_n\right)\right\}\left\{\hat{g}_{j,h}\left(\psi_{i-1}\right)-g_j\left(\psi_{i-1}\right)\right\}\omega\left(\psi_{i-1},\hat{\psi}_{i-1}\right)$$

where second term is equivalent to the average squared error of a one-step semiparametric estimation.

Similarly, the kernel-weighted LS estimator $\hat{\sigma}_h^2$ of σ^2 in (2.15) can also be rewritten as

$$\hat{\sigma}_h^2=\frac{1}{(T-1)}\sum_{i=2}^{T}\left\{x_i-\hat{\gamma}_h x_{i-1}-\hat{g}_{1,h}\left(\hat{\psi}_{i-1}\right)+\hat{\gamma}_h\hat{g}_{2,h}\left(\hat{\psi}_{i-1}\right)\right\}^2. \tag{4.10}$$

By decomposing (4.10) into

$$\frac{1}{T-1}\sum_{i=2}^{T}\left\{x_i-\left(\hat{\gamma}_h x_{i-1}+\hat{g}_{1,h}\left(\hat{\psi}_{i-1}\right)-\hat{\gamma}_h\hat{g}_{2,h}\left(\hat{\psi}_{i-1}\right)\right)\right\}^2$$
$$=\frac{1}{T-1}\sum_{i=2}^{T}\eta_i^2+\frac{1}{T-1}\sum_{i=2}^{T}\hat{z}_i^2\left(\gamma-\hat{\gamma}_h\right)$$
$$+\frac{1}{T-1}\sum_{i=2}^{T}\left\{\left[g_1\left(\psi_{i-1}\right)-\hat{g}_{1,h}\left(\hat{\psi}_{i-1}\right)\right]-\gamma\left[g_2\left(\psi_{i-1}\right)-\hat{g}_{2,h}\left(\hat{\psi}_{i-1}\right)\right]\right\}^2 \tag{4.11}$$
$$+\left(\gamma-\hat{\gamma}_h\right)\frac{2}{T-1}\sum_{i=2}^{T}\eta_i\hat{z}_i+\frac{2}{T-1}\sum_{i=2}^{T}\left\{\left[g_1\left(\psi_{i-1}\right)-\hat{g}_{1,h}\left(\hat{\psi}_{i-1}\right)\right]-\gamma\left[g_2\left(\psi_{i-1}\right)-\hat{g}_{2,h}\left(\hat{\psi}_{i-1}\right)\right]\right\}\eta_i$$
$$+\left(\gamma-\hat{\gamma}_h\right)\frac{2}{T-1}\sum_{i=2}^{T}\left\{\left[g_1\left(\psi_{i-1}\right)-\hat{g}_{1,h}\left(\hat{\psi}_{i-1}\right)\right]-\gamma\left[g_2\left(\psi_{i-1}\right)-\hat{g}_{2,h}\left(\hat{\psi}_{i-1}\right)\right]\right\}\hat{z}_i,$$

we can prove the following proposition.

Proposition 4b:

Assume that Assumptions 4a hold. Then the following holds uniformly over $h\in H_T$

$$\sqrt{T}\left(\hat{\sigma}_h^2-\sigma^2\right)\rightarrow N\left[0,Var\left(\eta_1^2\right)\right]. \tag{4.12}$$

In the discussion that follows, we will apply a cross-validation (CV) criterion to construct adaptive data-driven estimates for both γ and σ. However, in order to define the CV function, we must first introduce the following estimator.

For $1 \le n \le N = T-1$, let us define

$$\hat{g}_{j,n}(\hat{\psi}_n) = \hat{g}_{j,n}(\hat{\psi}_n, h) = \frac{1}{N-1}\sum_{s \ne n} \frac{K_h(\hat{\psi}_n - \hat{\psi}_s) x_{s+2-j}}{\hat{f}_{h,n}(\hat{\psi}_n)} \tag{4.13}$$

and

$$\hat{g}_{h,n}(\hat{\psi}_n) = \hat{g}_{1,n}(\hat{\psi}_n) - \gamma \hat{g}_{2,n}(\hat{\psi}_n) \tag{4.14}$$

where $\hat{f}_{h,n}(\hat{\psi}_n) = \frac{1}{N-1}\sum_{s \ne n} K_h(\hat{\psi}_n - \hat{\psi}_s)$. Similar to that of (2.14), the leave-out estimate $\tilde{\gamma}_h$ of γ can be founded by minimising

$$\sum_{n=1}^{N} \left\{ x_{n+1} - \gamma x_n - \hat{g}_{h,n}(\hat{\psi}_n) \right\}^2 . \tag{4.15}$$

The CV function for our current study can therefore be defined as

$$CV(h) = \frac{1}{N}\sum_{n=1}^{N} \eta_{n+1}^2 \omega(\psi_{i-1}, \hat{\psi}_{i-1}) + \bar{\mathfrak{D}}(h) - 2\mathfrak{C}(h) \tag{4.16}$$

where

$$\bar{\mathfrak{D}}(h) = \frac{1}{N}\sum_{n=1}^{N} \left\{ \hat{m}_{h,n}(\hat{v}_n) - m(v_n) \right\}^2 \omega(\hat{v}_n, v_n), \tag{4.17}$$

$$\mathfrak{C}(h) = \frac{1}{N}\sum_{n=1}^{N} \eta_{n+1} \left\{ \hat{m}_{h,n}(\hat{v}_n) - m(v_n) \right\} \omega(\hat{v}_n, v_n), \tag{4.18}$$

$$\hat{m}_{h,n}(\hat{v}_n) = \tilde{\gamma}_h x_n + \hat{g}_{1,n}(\hat{\psi}_n) - \tilde{\gamma}_h \hat{g}_{2,n}(\hat{\psi}_n), \tag{4.19}$$

and

$$m(v_n) = \gamma x_n + g_1(\psi_n) - \gamma g_2(\psi_n). \tag{4.20}$$

The cross-validation criterion consists of selecting the value $\hat{h}_C$ of h that achieves

$$CV(\hat{h}_C) = \inf_{h \in H_T} CV(h). \tag{4.21}$$

Furthermore, a data-driven selection $\hat{h}$ is asymptotically optimal if

$$\mathfrak{D}(\hat{h}) \big/ \inf_{h \in H_T} \mathfrak{D}(h) \to_p 1. \tag{4.22}$$

Proposition 4c:

Assume that Assumptions 4a hold. Then the data-driven bandwidth $\hat{h}_C$ is asymptotically optimal in accordance with (4.22).

To provide the proof of Proposition 4c, it is enough to show that

$$\mathfrak{D}(h) = \frac{1}{(T-1)} \sum_{i=2}^{T} \left[\left\{ \hat{\gamma}_h x_{i-1} + \hat{g}_{1,h}(\hat{\psi}_{i-1}) - \hat{\gamma}_h \hat{g}_{2,h}(\hat{\psi}_{i-1}) \right\} - \left\{ \gamma x_{i-1} + g_1(\psi_{i-1}) - \gamma g_2(\psi_{i-1}) \right\} \right]^2 \omega(\hat{\psi}_{i-1}, \psi_{i-1})$$
$$= \frac{C_1}{Th} + C_2 h^4 + o_P\left\{ \mathfrak{D}(h) \right\}, \tag{4.23}$$

where C_i $(i = 1, 2)$ are some positive constants, that

$$\sup_{h \in H_T} \frac{\left| \overline{\mathfrak{D}}(h) - \mathfrak{D}(h) \right|}{\mathfrak{D}(h)} = o_P(1) \tag{4.24}$$

and that

$$\sup_{h, h_1 \in H_T} \left| \left\{ \overline{\mathfrak{D}}(h) - \overline{\mathfrak{D}}(h_1) \right\} - \left\{ CV(h) - CV(h_1) \right\} \right| / \mathfrak{D}(h) = o_P(1). \tag{4.25}$$

The following section presents a small sample study for model (2.7)

5. Computational Aspects and Illustrative Examples

The aim of the following simulation experiments is to illustrate how well the above estimation procedure works numerically and practically. To demonstrate some degree of robustness, this paper considers a number of illustrative examples where, in each case, the random variable ε has either a Gammaa distribution with $\kappa = 2$ and $\beta = 0.5$, or a Weibull distribution with $\alpha = 3$ and $\beta = 1$. Also, in the analysis below, we use the quartic kernel function

$$K(u) = \begin{cases} (15/16)(1-u^2)^2 & \text{if } |u| \le 1 \\ 0 & \text{otherwise} \end{cases}. \tag{5.1}$$

and the weight function

$$w(s) = \begin{cases} 1 & \text{if } |s| \le 5 \\ 0 & \text{otherwise} \end{cases}. \tag{5.2}$$

The remaining of this section presents first descriptions of the illustrative examples considered in this paper, then Section 5b explains the computational steps taken in each case, including the first set of simulation results. Finally, Section 5c compares our simulation results to those of Allen et al. (2006).

5a. Illustrative Examples

Example 1: Mackey-Glass ACD (MG-ACD) Model

Our first example model stems from the well-known Mackey-Glass System.[7] In view of (2.7) a Mackey-Glass ACD model can be established by specifying

$$\gamma = 0.5 \text{ and } g(\psi) = 0.75\left(\frac{\psi}{1+\psi^2}\right). \tag{5.3}$$

Given the functional form of g, the fact that the process $\{\psi_i\}$ is strictly stationary follows from Theorem 3.1 of An and Huang (1996). Furthermore, Lemma 3.4.4 and Theorem 3.4.10 of Györfi et al (1989) suggest that the $\{\psi_i\}$ is β-mixing and therefore α-mixing. Finally, it follows from the definitions of K and w given in (5.1) and (5.2) above that all the conditions in Assumption 4a are satisfied.

Example 2: Logarithmic ACD (Log-ACD) Model

Similar to the GARCH, the ACD model specified in (2.7) requires additional restrictions to ensure the positivity of duration. To address this issue, Bauwens and Giot (2000) propose the Logarithmic ACD (Log-ACD) model as follows

$$x_i = \exp(\varphi_i)\varepsilon_i,\ \varphi_i = \omega + \sum_{j=1}^{p}\alpha_j x_{i-j} + \sum_{k=1}^{q}\beta_k\psi_{i-k}. \tag{5.4}$$

where $\{\varepsilon_i\} \sim$ i.i.d. with $E(\varepsilon_i) = v$. Let us also define

$$\exp(\phi_i) = v\exp(\varphi_i) \tag{5.5}$$

thereby (5.4) can now be rewritten as

$$x_i = \exp(\phi_i)\eta_i,\ \phi_i = \varpi + \sum_{j=1}^{p}\alpha_j \ln x_{i-j} + \sum_{k=1}^{q}\beta_j\phi_{i-j} \tag{5.6}$$

where $\varpi = \omega + \ln v$ and $\eta_i = \varepsilon_i/v$ such that $E(\eta_i) = 1$.

---

[7] See, for example, Nychka et al. (1992).

The so-called $\text{Log-ACD}(1,1)$ model can be written as

$$x_i = \exp(\phi_i)\eta_i,\ \phi_i = \varpi + \alpha \ln x_{i-1} + \beta\phi_{i-1} \tag{5.7}$$

Similar to its ACD(1,1) counterpart, the $\text{Log-ACD}(1,1)$ model can also be linearised to obtain

$$\ln x_i = \varpi + \alpha \ln x_{i-j} + \beta\phi_{i-1} + \mu_i \tag{5.8}$$

where $\mu_i = \ln \eta_i - 1$, so that $E(\mu_i) = 0$.

Below, we illustrate how well the SP-ACD model perform in the case where the data generating process for each of the realisations is given by the following $\text{Log-ACD}(1,1)$ model

$$x_i = \exp(\phi_i)\eta_i \text{ and } \phi_i = 0.01 + 0.2\ln x_{i-1} + 0.7\phi_{i-1}. \tag{5.9}$$

In Log-ACD studies, for example Bauwens and Giot (2000) and Allen et al. (2006), the parameters

$$\theta = (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q) \tag{5.10}$$

in (5.6) are estimated by the Maximum Likelihood (ML) method where the functional form of the likelihood function depends very much on the distribution of ε_i. If the distribution specified in the likelihood function is different from the true distribution of ε_i, then $\hat{\theta}$ is in fact the Quasi MLE (QMLE) of θ. Allen et al. (2006) empirically studied the finite sample properties of the MLE and QMLE as applied to the Log-ACD(1,1) model based on a variety of probability distributions, including the Weibull, the exponential, the generalised Gamma and the log-normal distribution, and found that, except for cases of the QMLE under the generalised Gamma distribution, the estimators are close to their true values and seems to be asymptotically normal. The problem with the generalised Gamma distribution may have been caused by difficulties in obtaining robust and accurate numerical derivatives of the likelihood functions for purposes of maximisation. Therefore, by assuming that the standardised duration ε has either the Weibull or the generalised Gamma distribution, our results in this paper can be directly compared to those of Allen et al. (2006).

5b. Computational Steps and Basic Simulation Results

The computational steps taken in order to obtain the results in Tables 1 to 4 below can be summarised as follows.

1) Compute

$$\mathfrak{D}(h)=\frac{1}{(T-1)}\sum_{i=2}^{T}\left[\left\{\hat{\gamma}_h x_{i-1}+\hat{g}_{1,h}\left(\hat{\psi}_{i-1}\right)-\hat{\gamma}_h\hat{g}_{2,h}\left(\hat{\psi}_{i-1}\right)\right\}-\left\{\gamma x_{i-1}+g_1\left(\psi_{i-1}\right)-\gamma g_2\left(\psi_{i-1}\right)\right\}\right]^2\omega\left(\hat{\psi}_{i-1},\psi_{i-1}\right)$$

and let $\hat{h}_D=\arg\min_{h\in H_T}$ where $H_T=\left[T^{-7/30},1.1T^{-1/6}\right]$.

2) Find the $\hat{h}=\hat{h}_C$ such that $\hat{h}_C=\arg\min_{h\in T} CV(h)$.

3) Under the cases where T=101, 201, 304 and 401, compute

   a) $\left|\hat{h}_C-\hat{h}_D\right|$,

   b) $\left|\hat{\gamma}_h-\gamma\right|$,

   c) $\left|\tilde{\gamma}_h-\gamma\right|$ and

   d) $ASE_1(h)$, $ASE_2(h)$, and $ASE_3(h)$

   where

$$ASE_1\left(\hat{h}_C\right)=\frac{1}{T-1}\sum_{i=2}^{T}\left\{\hat{g}^*_{\hat{h}_C,i}\left(\hat{\psi}_{i-1}\right)-\hat{g}^*_{\hat{h}_C,i}\left(\psi_{i-1}\right)\right\}^2, \tag{5.11}$$

$$ASE_2\left(\hat{h}_C\right)=\frac{1}{T-1}\sum_{i=2}^{T}\left\{\hat{g}^*_{\hat{h}_C,i}\left(\psi_{i-1}\right)-g\left(\psi_{i-1}\right)\right\}^2 \tag{5.12}$$

$$ASE_3\left(\hat{h}_C\right)=\frac{2}{T-1}\sum_{i=2}^{T}\left\{\hat{g}^*_{\hat{h}_C,i}\left(\hat{\psi}_{i-1}\right)-\hat{g}^*_{\hat{h}_C,i}\left(\psi_{i-1}\right)\right\}\left\{\hat{g}^*_{\hat{h}_C,i}\left(\psi_{i-1}\right)-g\left(\psi_{i-1}\right)\right\} \tag{5.13}$$

   and

$$ASE\left(\hat{h}_C\right)=ASE_1\left(\hat{h}_C\right)+ASE_2\left(\hat{h}_C\right)+ASE_3\left(\hat{h}_C\right). \tag{5.14}$$

All simulations were performed in S-plus. The means of the results for all four cases, namely the Weibull MG-ACD (WMG-ACD), the Gamma MG-ACD (GMG-ACD), the Weibull Log-ACD (WL-ACD) and the Gamma Log-ACD (GL-ACD) models, are tabulated in Tables 1 to 4, respectively. Note also that in these tables N, R and M denote $T-1$, the number of replications and the number of basic iterations, respectively. We will now discuss a number of important findings.

Firstly, the simulation results in Tables 1 to 4 show that in all four cases the absolute error $\left|\hat{h}_C - \hat{h}_D\right|$ has the tendency to converge to zero as N increases. It should be noted, however, that $\hat{h}_C$ and $\hat{h}_D$ shown here are those of the final estimation step only. Although the results are not reported here, we also considered $\hat{h}_C$ and $\hat{h}_D$ at each of the mth iteration and found that the absolute error $\left|\hat{h}_C - \hat{h}_D\right|$ has the tendency to converge to zero in all cases.

Secondly, our estimation method is able to provide estimates for the parameter γ with comparable degree of accuracy to those of a one-step partially linear autoregressive estimation reported, for example, in Gao and Yee (2000). In all four cases, the absolute errors $|\hat{\gamma}_h - \gamma|$ and $|\tilde{\gamma}_h - \gamma|$ have the tendency to converge to zero as $\mathrm{N} \to \infty$ at a similar rate as those reported in Tables 1 and 2 of Gao and Yee (2000). These results are quite stable and are not significantly affected by increases in the number of replications, R. However, it is interesting to report that our estimation method seems to perform better, with respect to $|\hat{\gamma}_h - \gamma|$ and $|\tilde{\gamma}_h - \gamma|$, at a smaller number of basic iteration, M, when applied to WMG-ACD and WL-ACD models, while performs better at a larger number of M when applied to GMG-ACD and GL-ACD. Furthermore, switching from the Weibull to the Gamma distributed standardised duration seems to have affected the results, with respect to both $|\hat{\gamma}_h - \gamma|$ $\left(\text{and } |\tilde{\gamma}_h - \gamma|\right)$ and $ASE\left(\hat{h}_C\right)$ (discussed below) substantially. In all aspects, $|\hat{\gamma}_h - \gamma|$ and $|\tilde{\gamma}_h - \gamma|$ of the GMG-ACD model are much larger than those of the WMG-ACD at a smaller number of observation, while the results becomes more comparable as N increases.

It is obvious from (5.11) to (5.14) that in our analysis the average square error is decomposed into

$$\begin{aligned} ASE\left(\hat{h}_C\right) &= \frac{1}{T-1}\sum_{i=2}^{T}\left\{\hat{g}^*_{\hat{h}_C,i}\left(\hat{\psi}_{i-1}\right) - \hat{g}^*_{\hat{h}_C,i}\left(\psi_{i-1}\right)\right\}^2 + \frac{1}{T-1}\sum_{i=2}^{T}\left\{\hat{g}^*_{\hat{h}_C,i}\left(\psi_{i-1}\right) - g\left(\psi_{i-1}\right)\right\}^2 \\ &\quad + \frac{2}{T-1}\sum_{i=2}^{T}\left\{\hat{g}^*_{\hat{h}_C,i}\left(\hat{\psi}_{i-1}\right) - \hat{g}^*_{\hat{h}_C,i}\left(\psi_{i-1}\right)\right\}\left\{\hat{g}^*_{\hat{h}_C,i}\left(\psi_{i-1}\right) - g\left(\psi_{i-1}\right)\right\} \\ &\equiv ASE_1\left(\hat{h}_C\right) + ASE_2\left(\hat{h}_C\right) + ASE_3\left(\hat{h}_C\right). \end{aligned} \tag{5.15}$$

Observe that $ASE_2\left(\hat{h}_C\right)$ in this case is equivalent to those of a one-step partially linear study. Therefore, it is common to see that our results of $ASE_2\left(\hat{h}_C\right)$ for the MG-ACD model in Tables 1 and 2 are quite comparable to those reported in Tables 1 and 2 of Gao and Yee (2000). However, notice in Tables 3 and 4 that $ASE_2\left(\hat{h}_C\right)$ for the Log-ACD models are relatively large compared to those of the MG-ACD models. This should not surprising because given the linear nature of the Log-ACD we would normally expect the above partially linear regression to perform better with the MG style model.

We will now turn our attention to the results of $ASE_1\left(\hat{h}_C\right)$, which represents estimation errors due to the fact that the conditional durations employed in our analysis are estimates. The simulation results in Table 1 to 4 indicate that $ASE_1\left(\hat{h}_C\right)$ in all cases, with only exception to that of GMG-ACD model, are significantly smaller than $ASE_2\left(\hat{h}_C\right)$ and have the tendency to converge quite fast to zero as N increases. However, $ASE_1\left(\hat{h}_C\right)$ of the GMG-ACD model are relatively large compared to those of its WMG-ACD counterpart in Table1. The highest $ASE_1\left(\hat{h}_C\right)$ in Table 2 was 0.0266 compared to only 0.0011 in Table 1. In addition, it is apparent from Table 2 that in this case $ASE_1\left(\hat{h}_C\right)$ have less tendency to converge to zero. In order to investigate these problems in more details, we will first write $ASE_1\left(\hat{h}_C\right)$ as

$$\begin{aligned}
ASE_1\left(\hat{h}_C\right) &= \frac{1}{T-1}\sum_{i=2}^{T}\left\{\hat{g}_{1,h}\left(\hat{\psi}_{i-1}\right)-\hat{g}_{1,h}\left(\psi_{i-1}\right)\right\}^2\omega\left(\hat{\psi}_{i-1},\psi_{i-1}\right) \\
&+\frac{\hat{\gamma}_{h,\hat{\psi}}^2}{T-1}\sum_{i=2}^{T}\left\{\hat{g}_{2,h}\left(\hat{\psi}_{i-1}\right)-\hat{g}_{2,h}\left(\psi_{i-1}\right)\right\}^2\omega\left(\hat{\psi}_{i-1},\psi_{i-1}\right) \\
&+\frac{\left\{\left(\hat{\gamma}_{h,\psi}-\gamma\right)+\left(\gamma-\hat{\gamma}_{h,\hat{\psi}}\right)\right\}^2}{T-1}\sum_{i=2}^{T}\hat{g}_{2,h}^2\left(\psi_{i-1}\right)\omega\left(\hat{\psi}_{i-1},\psi_{i-1}\right) \\
&+\frac{2}{T-1}\sum_{i=2}^{T}\left\{\hat{g}_{1,h}\left(\hat{\psi}_{i-1}\right)-\hat{g}_{1,h}\left(\psi_{i-1}\right)\right\}\left\{\hat{g}_{2,h}\left(\hat{\psi}_{i-1}\right)-\hat{g}_{2,h}\left(\psi_{i-1}\right)\right\}\omega\left(\hat{\psi}_{i-1},\psi_{i-1}\right) \\
&+\frac{2}{T-1}\left\{\left(\hat{\gamma}_{h,\psi}-\gamma\right)+\left(\gamma-\hat{\gamma}_{h,\hat{\psi}}\right)\right\}\sum_{i=2}^{T}\left\{\hat{g}_{1,h}\left(\hat{\psi}_{i-1}\right)-\hat{g}_{1,h}\left(\psi_{i-1}\right)\right\}\hat{g}_{2,h}\left(\psi_{i-1}\right)\omega\left(\hat{\psi}_{i-1},\psi_{i-1}\right) \\
&+\frac{2}{T-1}\hat{\gamma}_{h,\hat{\psi}}\sum_{i=2}^{T}\left\{\hat{g}_{2,h}\left(\hat{\psi}_{i-1}\right)-\hat{g}_{2,h}\left(\psi_{i-1}\right)\right\}\hat{g}_{2,h}\left(\psi_{i-1}\right)\omega\left(\hat{\psi}_{i-1},\psi_{i-1}\right)
\end{aligned} \tag{5.16}$$

where $\hat{\gamma}_{\psi}$ and $\hat{\gamma}_{\hat{\psi}}$ are the estimators in (2.14) and (4.1), respectively. Other things else being equal, it is obvious from (5.16) that for $ASE_1\left(\hat{h}_C\right)$ to be converging to a smaller value, both $\hat{\gamma}_{\psi}$ and $\hat{\gamma}_{\hat{\psi}}$ must be converging to γ. We have already mentioned earlier that $\left|\hat{\gamma}_{\hat{\psi}}-\gamma\right|$ are much larger and have the tendency to converge to zero at a much slower rate here when compared to those of its WMG-ACD model counterpart. Furthermore, our investigation has found that at $\mathrm{N}=100, 200, 300,$ and 400, $\left|\hat{\gamma}_{\psi}-\gamma\right|$ of the GMG-ACD model are 0.1685, 0.1396, 0.1333 and 0.1292, respectively. These are relatively large and slow convergence compared to those of the WMG-ACD model of 0.0911, 0.0769, 0.0710 and 0.0656.

5c. SP-ACD, MLE and QMLE Comparison

In this section, we compare our simulation results of the WL-ACD and the GL-ACD model with those of Allen et al. (2006). While, the first column of Tables 5 and 6 below presents descriptive statistics of the estimates $\hat{\gamma}_{\hat{h}_C}$, the second, the third and the fourth show those of Allen et al. (2006) with respect to the MLE, the best and the worst of the QMLE, respectively. To obtain the results in Tables 5 and 6 below, we follow similar computational steps as those discussed in Section 5b under cases where $\mathrm{N}=500$ and $1{,}000$, $\mathrm{R}=500$ and $\mathrm{M}=3$. Our experience shows that changes in R and M do not have significant effect on the results. We will now discuss some important findings.

Although, the results are not shown on the tables, our investigation has found that at $\mathrm{N}=1{,}000$ the absolute error about γ, $\left|\hat{\gamma}_{\hat{h}_C}-\gamma\right|$, declined to 0.0477 and 0.0396 for cases of the WL-ACD and the GL-ACD, respectively. Furthermore, simulation results in Tables 5 and 6 show that on average our estimation method tends to overestimate the value of the parameter γ. Though increasing the number of observation from 500 to 1,000 affects accuracy only slightly, it helps improve estimation precision significantly as evidenced by the decline in the estimation standard deviation from 0.057 to 0.040 and from 0.062 to 0.037 for the WL-ACD and the GL-ACD models, respectively. Also, as N increases the skewness and kurtosis of our estimates tend to those of normal distribution, i.e. zero and three, respectively.

Secondly, in all four cases, the ML estimation seems to provide the most accurate estimates among the three methods. However, the MLE for the GL-ACD model seems to be leptokurtic as can be evidenced by a relatively high kurtosis of 4.849 and 5.670 at $N = 500$ and $1,000$, respectively. Unlike the QML, our estimation method seems to have suffered from a similar problem, but with much less extent.

Thirdly, the simulation results in Tables 5 and 6 indicate that, where sample size is small and the true distribution of the standardised durations is not known, our estimation method should be preferred to the QML estimation. If the Weibull distribution is incorrectly assumed when the true DGP is in fact GL-ACD process, the QML would have substantially overestimated the true value of the parameter. Our method, on the other hand, is more robust across the distribution assumed and, hence, should be more reliable.

6. Discussion

We proposed in this paper the Semiparametric ACD (SP-ACD) model. To address the problem about the unobservability of the conditional durations in practice, the current paper devised an iterative algorithm to estimate the unknown conditional duration process. Mathematical analysis about the theoretical justification of the estimation algorithm is also provided. Although, our discussion went on without giving detailed consideration about the asymptotic theory of the semiparametric and the adaptive data-driven estimators in this case, our simulation analysis have shown promising results in the sense that our model seems to have satisfactory asymptotic characters, while its statistical performance is also quite robust across data generating processes and assumptions about the probability distribution of the standardised durations. Apart from the model's asymptotic theory, our future study will also involve investigation on some other distribution assumptions to provide more proofs of robustness and also further generalisation of the SP-ACD model, perhaps to also incorporate such a model as (2.16).

Table 1: WMG-ACD Model

N	100				200			
R	100		500		100		500	
M	3	8	3	8	3	8	3	8
$\left\|\hat{h}_C - \hat{h}_D\right\|$	0.0123	0.1203	0.1201	0.1200	0.1000	0.0980	0.0855	0.0824
$\left\|\hat{\gamma}_{\hat{h}_C} - \gamma\right\|$	0.0774	0.0770	0.0809	0.0814	0.0706	0.0706	0.0636	0.0633
$\left\|\tilde{\gamma}_{\hat{h}_C} - \gamma\right\|$	0.0773	0.0769	0.0809	0.0813	0.0706	0.0706	0.0636	0.0634
$ASE1\left(\hat{h}_C\right)$	0.0009	0.0011	0.0008	0.0009	0.0006	0.0005	0.0005	0.0004
$ASE2\left(\hat{h}_C\right)$	0.0065	0.0070	0.0066	0.0065	0.0046	0.0046	0.0045	0.0045
$ASE3\left(\hat{h}_C\right)$	(0.0019)	(0.0026)	(0.0019)	(0.0019)	(0.0018)	(0.0018)	(0.0016)	(0.0016)
$ASE\left(\hat{h}_C\right)$	0.0056	0.0054	0.0055	0.0055	0.0033	0.0033	0.0035	0.0034

N	300				400			
R	100		500		100		500	
M	3	8	3	8	3	8	3	8
$\left\|\hat{h}_C - \hat{h}_D\right\|$	0.0670	0.0620	0.0638	0.0630	0.0468	0.0466	0.0490	0.0480
$\left\|\hat{\gamma}_{\hat{h}_C} - \gamma\right\|$	0.0602	0.0608	0.0561	0.0561	0.0539	0.0540	0.0513	0.0511
$\left\|\tilde{\gamma}_{\hat{h}_C} - \gamma\right\|$	0.0602	0.0608	0.0561	0.0561	0.0539	0.0540	0.0513	0.0511
$ASE1\left(\hat{h}_C\right)$	0.0006	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004
$ASE2\left(\hat{h}_C\right)$	0.0038	0.0038	0.0041	0.0041	0.0041	0.0041	0.0036	0.0036
$ASE3\left(\hat{h}_C\right)$	(0.0014)	(0.0013)	(0.0016)	(0.0016)	(0.0015)	(0.0015)	(0.0014)	(0.0014)
$ASE\left(\hat{h}_C\right)$	0.0029	0.0029	0.0029	0.0029	0.0030	0.0029	0.0027	0.0026

Table 2: GMG-ACD Model

N	100				200			
R	100		500		100		500	
M	3	8	3	8	3	8	3	8
$\left\|\hat{h}_C - \hat{h}_D\right\|$	0.0737	0.0858	0.0857	0.0910	0.0673	0.0655	0.0684	0.0668
$\left\|\hat{\gamma}_{\hat{h}_C} - \gamma\right\|$	0.1320	0.1300	0.1280	0.1235	0.1057	0.0959	0.1111	0.1017
$\left\|\tilde{\gamma}_{\hat{h}_C} - \gamma\right\|$	0.1321	0.1300	0.1281	0.1236	0.1058	0.0960	0.1112	0.1018
$ASE1\left(\hat{h}_C\right)$	0.0099	0.0115	0.0100	0.0127	0.0179	0.0266	0.0148	0.0199
$ASE2\left(\hat{h}_C\right)$	0.0152	0.0150	0.0117	0.0117	0.0096	0.0092	0.0111	0.0118
$ASE3\left(\hat{h}_C\right)$	(0.0077)	(0.0084)	(0.0061)	(0.0064)	(0.0097)	(0.0125)	(0.0084)	(0.0094)
$ASE\left(\hat{h}_C\right)$	0.0173	0.0181	0.0156	0.0184	0.0177	0.0238	0.0175	0.0217

N	300				400			
R	100		500		100		500	
M	3	8	3	8	3	8	3	8
$\left\|\hat{h}_C - \hat{h}_D\right\|$	0.0586	0.0574	0.0577	0.0609	0.0575	0.0594	0.0507	0.0523
$\left\|\hat{\gamma}_{\hat{h}_C} - \gamma\right\|$	0.0938	0.0888	0.1083	0.0991	0.08395	0.0781	0.0818	0.0735
$\left\|\tilde{\gamma}_{\hat{h}_C} - \gamma\right\|$	0.0939	0.0888	0.1084	0.0992	0.08398	0.0782	0.0820	0.0736
$ASE1\left(\hat{h}_C\right)$	0.0137	0.0170	0.0155	0.0211	0.0191	0.0195	0.0206	0.0206
$ASE2\left(\hat{h}_C\right)$	0.0087	0.0087	0.0078	0.0100	0.0079	0.0079	0.0044	0.0044
$ASE3\left(\hat{h}_C\right)$	(0.0082)	(0.0097)	(0.0087)	(0.0098)	(0.0079)	(0.0088)	(0.0102)	(0.0114)
$ASE\left(\hat{h}_C\right)$	0.0142	0.0160	0.0146	0.0213	0.0191	0.0196	0.0148	0.0136

Table 3: WL-ACD Model

N	100				200			
R	100		500		100		500	
M	3	8	3	8	3	8	3	8
$\left\|\hat{h}_C - \hat{h}_D\right\|$	0.1521	0.1506	0.1658	0.1718	0.1233	0.1277	0.1277	0.1279
$\left\|\hat{\gamma}_{\hat{h}_C} - \gamma\right\|$	0.0981	0.1031	0.1049	0.1093	0.0790	0.0828	0.0800	0.0832
$\left\|\tilde{\gamma}_{\hat{h}_C} - \gamma\right\|$	0.0975	0.1027	0.1037	0.1085	0.0789	0.0828	0.0800	0.0832
$ASE1\left(\hat{h}_C\right)$	0.0018	0.0017	0.0020	0.0018	0.0007	0.0006	0.0004	0.0006
$ASE2\left(\hat{h}_C\right)$	0.0113	0.0113	0.0152	0.0152	0.0144	0.0144	0.0120	0.0120
$ASE3\left(\hat{h}_C\right)$	0.0028	0.0026	0.0041	0.0036	0.0027	0.0025	0.0028	0.0026
$ASE\left(\hat{h}_C\right)$	0.0160	0.0157	0.0214	0.0207	0.0179	0.0176	0.0135	0.0158

N	300				400			
R	100		500		100		500	
M	3	8	3	8	3	8	3	8
$\left\|\hat{h}_C - \hat{h}_D\right\|$	0.1033	0.0983	0.0979	0.0933	0.1151	0.1127	0.1008	0.0980
$\left\|\hat{\gamma}_{\hat{h}_C} - \gamma\right\|$	0.0688	0.0712	0.0702	0.0727	0.0621	0.0647	0.0577	0.0597
$\left\|\tilde{\gamma}_{\hat{h}_C} - \gamma\right\|$	0.0688	0.0713	0.0703	0.0728	0.0622	0.06491	0.0577	0.0598
$ASE1\left(\hat{h}_C\right)$	0.0006	0.0005	0.0007	0.0007	0.0015	0.0005	0.0007	0.0005
$ASE2\left(\hat{h}_C\right)$	0.0117	0.0120	0.0131	0.0131	0.0125	0.0125	0.0116	0.0116
$ASE3\left(\hat{h}_C\right)$	0.0035	0.0030	0.0034	0.0032	0.0011	0.0016	0.0023	0.0022
$ASE\left(\hat{h}_C\right)$	0.0158	0.0156	0.0173	0.0171	0.0152	0.0147	0.01468	0.0144

Table 4: GL-ACD Model

N	100				200			
R	100		500		100		500	
M	3	8	3	8	3	8	3	8
$\left\lvert \hat{h}_C - \hat{h}_D \right\rvert$	0.2621	0.2540	0.2682	0.2586	0.1482	0.1430	0.1299	0.1217
$\left\lvert \hat{\gamma}_{\hat{h}_C} - \gamma \right\rvert$	0.0970	0.09637	0.0972	0.0970	0.0776	0.0785	0.0709	0.0733
$\left\lvert \tilde{\gamma}_{\hat{h}_C} - \gamma \right\rvert$	0.0973	0.0969	0.0968	0.0969	0.0778	0.0786	0.0710	0.0733
$ASE1\left(\hat{h}_C\right)$	0.0064	0.0049	0.0060	0.0053	0.0023	0.0019	0.0034	0.0036
$ASE2\left(\hat{h}_C\right)$	0.0430	0.0430	0.0351	0.0351	0.0232	0.0329	0.0348	0.0348
$ASE3\left(\hat{h}_C\right)$	0.0130	0.0107	0.0091	0.00836	0.0097	0.0098	0.0104	0.0093
$ASE\left(\hat{h}_C\right)$	0.0626	0.0587	0.0503	0.0482	0.0450	0.0446	0.0486	0.0477

N	300				400			
R	100		500		100		500	
M	3	8	3	8	3	8	3	8
$\left\lvert \hat{h}_C - \hat{h}_D \right\rvert$	0.1105	0.1039	0.0969	0.0939	0.0760	0.0710	0.0810	0.0523
$\left\lvert \hat{\gamma}_{\hat{h}_C} - \gamma \right\rvert$	0.0685	0.0656	0.0593	0.0600	0.0584	0.0591	0.0544	0.0935
$\left\lvert \tilde{\gamma}_{\hat{h}_C} - \gamma \right\rvert$	0.0654	0.0657	0.0594	0.0602	0.0586	0.0593	0.0545	0.0935
$ASE1\left(\hat{h}_C\right)$	0.0030	0.0025	0.0028	0.0025	0.0021	0.0017	0.0028	0.0094
$ASE2\left(\hat{h}_C\right)$	0.0590	0.0509	0.0405	0.0405	0.0312	0.0312	0.0420	0.0262
$ASE3\left(\hat{h}_C\right)$	(0.0004)	0.0001	0.0066	0.0061	0.0095	0.0086	0.0147	(0.0114)
$ASE\left(\hat{h}_C\right)$	0.0530	0.0536	0.0500	0.0492	0.0429	0.0416	0.0595	0.0242

Table 5: WL-ACD Model

N = 500 R = 500	SP-ACD(1,1)	ML Weibull Dist	Best of QML Exponential Dist	Worst of QML Gammaa Dist
Mean	0.239	0.201	0.199	0.199
Maximum	0.466	0.313	0.318	0.677
Minimum	0.061	0.099	0.090	(0.115)
Std. Dev.	0.057	0.028	0.028	0.097
Skewness	0.325	0.068	0.066	0.358
Kurtosis	3.290	3.188	3.192	4.255

N = 1,000 R = 500	SP-ACD(1,1)	ML Weibull Dist	Best of QML Exponential Dist	Worst of QML Gammaa Dist
Mean	0.237	0.200	0.199	0.198
Maximum	0.378	0.276	0.286	0.729
Minimum	0.142	0.126	0.127	(0.095)
Std. Dev.	0.040	0.019	0.019	0.077
Skewness	0.295	0.077	0.099	0.486
Kurtosis	3.262	3.119	3.031	6.375

Table 6: GL-ACD Model

N = 500 R = 500	SP-ACD(1,1)	ML Gammaa Dist	Best of QML Exponential Dist	Worst of QML Weibull Dist
Mean	0.233	0.201	0.197	0.252
Maximum	0.443	0.716	0.328	0.399
Minimum	0.040	(0.100)	0.097	0.121
Std. Dev.	0.062	0.078	0.031	0.038
Skewness	0.230	0.737	0.075	0.256
Kurtosis	3.679	4.849	3.170	3.190

N = 1,000 R = 500	SP-ACD(1,1)	ML Gammaa Dist	Best of QML Exponential Dist	Worst of QML Weibull Dist
Mean	0.232	0.202	0.199	0.236
Maximum	0.344	0.630	0.282	0.329
Minimum	0.113	0.037	0.129	0.136
Std. Dev.	0.037	0.057	0.022	0.027
Skewness	0.100	0.826	(0.021)	0.148
Kurtosis	3.350	5.670	3.122	3.062

Appendix

Proof of Theorem 3a

The first L_2 estimation errors in (3.9) for the mth iteration can be formally written as

$$\left\|\hat{\psi}_{1,n,m}-\overline{\psi}_{1,n,m}\right\|_2^2, \left\|\hat{\psi}_{2,n,m}-\overline{\psi}_{2,n,m}\right\|_2^2, \ldots, \left\|\hat{\psi}_{T,n,m}-\overline{\psi}_{T,n,m}\right\|_2^2. \tag{3a.1}$$

where

$$\begin{aligned} \left\|\hat{\psi}_{1,n,m}-\overline{\psi}_{1,n,m}\right\|_2^2 &= E\left[\hat{\psi}_{1,n,m}-\overline{\psi}_{1,n,m}\right]^2 \\ \left\|\hat{\psi}_{2,n,m}-\overline{\psi}_{2,n,m}\right\|_2^2 &= E\left[\hat{\psi}_{2,n,m}-\overline{\psi}_{2,n,m}\right]^2 \\ \vdots \qquad & \qquad \vdots \\ \left\|\hat{\psi}_{T,n,m}-\overline{\psi}_{T,n,m}\right\|_2^2 &= E\left[\hat{\psi}_{T,n,m}-\overline{\psi}_{T,n,m}\right]^2. \end{aligned} \tag{3a.2}$$

Observe that, in (3.14), $\Delta_{1\kappa}^{(L_2)}$ defines the maximal errors corresponding to a particular pair of i and n across κ where $1 \le \kappa \le m-1$. Let now extend this definition to $1 \le m \le M$ such that

$$\Delta_{1m}^{(L_2)} =: \sup_{m\ge 1}\left\|\hat{\psi}_{i,m}-\overline{\psi}_{i,m}\right\|_2 =: \sup_{m\ge 1}\left\|\left\{\hat{\gamma}_h x_{i-1}+\hat{g}_h\left(\hat{\psi}_{i-1,m-1}\right)\right\}-\left\{\gamma x_{i-1}+g\left(\hat{\psi}_{i-1,m-1}\right)\right\}\right\|_2 \tag{3a.3}$$

where $\hat{\psi}_{i,m}=\left(\hat{\psi}_{i,1,m},\ldots,\hat{\psi}_{i,N,m}\right)^\tau$, $\overline{\psi}_{i,m}=\left(\overline{\psi}_{i,1,m},\ldots,\overline{\psi}_{i,N,m}\right)^\tau$ and $i=1,\ldots,T$. Furthermore, the asymptotic results in Section 4 suggest that we have

$$\frac{1}{(T-1)}\sum_{i=2}^{T}\left(\Delta_{1m}^{(L_2)}\right)^2 = \frac{C_1}{(T-1)h}+C_2h^4+o_P\left\{\frac{1}{(T-1)h}+h^4\right\} \tag{3a.4}$$

uniformly over $h \in H_T$. Therefore, using results in (3a.1) to (3a.4) and the fact that

$$\frac{1}{(T-1)}\sum_{i=2}^{T}\Delta_{1m}^{(L_2)} = \Delta_{1m}^{(L_2)} = \left(\frac{1}{(T-1)}\sum_{i=2}^{T}\left(\Delta_{1m}^{(L_2)}\right)^2\right)^{1/2} \tag{3a.5}$$

we conclude that

$$\Delta_{1m}^{(L_2)} = \left(\frac{C_1}{(T-1)h}+C_2h^4+o_P\left\{\frac{1}{(T-1)h}+h^4\right\}\right)^{1/2} \tag{3a.6}$$

uniformly over $h \in H_T$

Proof of Theorem 3b:

The definition in (3.10) suggests that we have

$$\begin{aligned} E\left|\psi_{i,n,m}-\psi_i\right|^2 &= E\left|\left\{g_1\left(\psi_{i-1,n,m-1}\right)-g_1\left(\psi_{i-1}\right)\right\}-\gamma\left\{g_2\left(\psi_{i-1,n,m-1}\right)-g_2\left(\psi_{i-1}\right)\right\}\right|^2 \\ &\le E\left\{\left|g_1\left(\psi_{i-1,n,m-1}\right)-g_1\left(\psi_{i-1}\right)\right|+|\gamma|\left|g_2\left(\psi_{i-1,n,m-1}\right)-g_2\left(\psi_{i-1}\right)\right|\right\}^2 \end{aligned} \tag{3b.1}$$

where

$$\begin{aligned} &E\left\{\left|g_1\left(\psi_{i-1,n,m-1}\right)-g_1\left(\psi_{i-1}\right)\right|+|\gamma|\left|g_2\left(\psi_{i-1,n,m-1}\right)-g_2\left(\psi_{i-1}\right)\right|\right\}^2 \\ &= E\left|g_1\left(\psi_{i-1,n,m-1}\right)-g_1\left(\psi_{i-1}\right)\right|^2 \\ &\quad +\gamma^2 E\left|g_2\left(\psi_{i-1,n,m-1}\right)-g_2\left(\psi_{i-1}\right)\right|^2 \\ &\quad +|\gamma| E\left|g_1\left(\psi_{i-1,n,m-1}\right)-g_1\left(\psi_{i-1}\right)\right|\left|g_2\left(\psi_{i-1,n,m-1}\right)-g_2\left(\psi_{i-1}\right)\right|. \end{aligned} \tag{3b.3}$$

We will need the following results in order to provide proof of Theorem 3b. Observe that by Assumption 3a and the Cauchy-Schwarz Inequality

$$\begin{aligned} E\left|g_j\left(\psi_{i-1,n,m-1}\right)-g_j\left(\psi_{i-1}\right)\right|^2 &\le E\left[\varphi_j^2\left(\psi_{i-1}\right)\left|\psi_{i-1,n,m-1}-\psi_{i-1}\right|^2\right] \\ &\le E\left[\varphi_j^2\left(\psi_{i-1}\right)\right]E\left|\psi_{i-1,n,m-1}-\psi_{i-1}\right|^2 \end{aligned} \tag{3b.3}$$

and, similarly,

$$\begin{aligned} E\left|g_1\left(\psi_{i-1,n,m-1}\right)-g_1\left(\psi_{i-1}\right)\right|\left|g_2\left(\psi_{i-1,n,m-1}\right)-g_2\left(\psi_{i-1}\right)\right| &\le \left\{E\left[\varphi_1^2\left(\psi_{i-1}\right)\right]E\left[\varphi_2^2\left(\psi_{i-1}\right)\right]\right\}^{1/2} \\ &\quad \cdot E\left|\psi_{i-1,n,m-1}-\psi_{i-1}\right|^2. \end{aligned} \tag{3b.4}$$

The above results suggest that we bound the quantity of interest in (3b.1) as follows

$$\begin{aligned} E\left|\psi_{i,n,m}-\psi_i\right|^2 &\le E\left[\varphi_1^2\left(\psi_{i-1}\right)\right]E\left|\psi_{i-1,n,m-1}-\psi_{i-1}\right|^2+\gamma^2 E\left[\varphi_2^2\left(\psi_{i-1}\right)\right]E\left|\psi_{i-1,n,m-1}-\psi_{i-1}\right|^2 \\ &\quad +2|\gamma|\left(E\left[\varphi_1^2\left(\psi_{i-1}\right)\right]E\left[\varphi_2^2\left(\psi_{i-1}\right)\right]\right)^{1/2}E\left|\psi_{i-1,n,m-1}-\psi_{i-1}\right|^2, \end{aligned} \tag{3b.5}$$

which can be conveniently simplified to

$$E\left|\psi_{i,n,m}-\psi_i\right|^2 \le C_{i-1}E\left|\psi_{i-1,n,m-1}-\psi_{i-1}\right|^2 \tag{3b.6}$$

where $C_{i-1}=E\left[\varphi_1^2\left(\psi_{i-1}\right)\right]+\gamma^2 E\left[\varphi_2^2\left(\psi_{i-1}\right)\right]+2|\gamma|\left(E\left[\varphi_1^2\left(\psi_{i-1}\right)\right]E\left[\varphi_2^2\left(\psi_{i-1}\right)\right]\right)^{1/2}$.

Equivalently, by taking square root on both sides of (3b.6), and referring to the norms of the vectors, the inequality could be written as

$$\left\|\psi_{i,n,m}-\psi_i\right\|_2 \le c_{i-1}\left\|\psi_{i-1,n,m-1}-\psi_{i-1}\right\|_2 \tag{3b.7}$$

where

$$\begin{aligned}\left(C_{i-1}\right)^{1/2} \le c_{i-1} &= \left\|\varphi_1\left(\psi_{i-1}\right)\right\|_2 + \left\|\varphi_2\left(\psi_{i-1}\right)\right\|_2 + 2\left(\left\|\varphi_1\left(\psi_{i-1}\right)\right\|_2 \left\|\varphi_2\left(\psi_{i-1}\right)\right\|_2\right)^{1/2} \\ &= \left(\left(\left\|\varphi_1\left(\psi_{i-1}\right)\right\|_2\right)^{1/2} + \left(\left\|\varphi_2\left(\psi_{i-1}\right)\right\|_2\right)^{1/2}\right)^2\end{aligned} \tag{3b.8}$$

thereby the inequality in (3b.8) can be verified by the Triangle Inequality and the fact that here the parameter γ takes on the maximum value, given the stationarity assumption on x_i, of unity.

Furthermore, performing backward iteration on (3b.7) gives rise to

$$\left\|\psi_{i,n,m}-\psi_i\right\|_2 \le c_{i-1}c_{i-2}\dots c_{i-m}\left\|\psi_{i-m,n,0}-\psi_{i-m}\right\|_2 \tag{3b.9}$$

where

$$c_{i-s} = \left(\left(\left\|\varphi_1\left(\psi_{i-1}\right)\right\|_2\right)^{1/2} + \left(\left\|\varphi_2\left(\psi_{i-1}\right)\right\|_2\right)^{1/2}\right)^2$$

for $s=1,\dots,m$. Finally, by using the second part of Assumption 3a, Assumption 3b and the Minkowski's Inequality for an L_2 would then lead to

$$\left\|\psi_{i,n,m}-\psi_i\right\|_2 \le G^m\left(V_1+V_2\right).$$

Proof of Theorem 3c:

First, observe that we have from (3.10) and the Triangle Inequality

$$\begin{aligned}E\left|\overline{\psi}_{i,n,m}-\psi_{i,n,m}\right|^2 &= E\left|\left\{g_1\left(\hat{\psi}_{i-1,n,m-1}\right)-g_1\left(\psi_{i-1,n,m-1}\right)\right\}-\gamma\left\{g_2\left(\hat{\psi}_{i-1,n,m-1}\right)-g_2\left(\psi_{i-1,n,m-1}\right)\right\}\right|^2 \\ &\le E\left\{\left|g_1\left(\hat{\psi}_{i-1,n,m-1}\right)-g_1\left(\psi_{i-1,n,m-1}\right)\right|+\left|\gamma\right|\left|g_2\left(\hat{\psi}_{i-1,n,m-1}\right)-g_2\left(\psi_{i-1,n,m-1}\right)\right|\right\}^2\end{aligned} \tag{3c.1}$$

where

$$\begin{aligned}&E\left\{\left|g_1\left(\hat{\psi}_{i-1,n,m-1}\right)-g_1\left(\psi_{i-1,n,m-1}\right)\right|+\left|\gamma\right|\left|g_2\left(\hat{\psi}_{i-1,n,m-1}\right)-g_2\left(\psi_{i-1,n,m-1}\right)\right|\right\}^2 \\ &\quad= E\left|g_1\left(\hat{\psi}_{i-1,n,m-1}\right)-g_1\left(\psi_{i-1,n,m-1}\right)\right|^2+\gamma^2 E\left|g_2\left(\hat{\psi}_{i-1,n,m-1}\right)-g_2\left(\psi_{i-1,n,m-1}\right)\right|^2 \\ &\quad+2\left|\gamma\right|E\left|g_1\left(\hat{\psi}_{i-1,n,m-1}\right)-g_1\left(\psi_{i-1,n,m-1}\right)\right|\left|g_2\left(\hat{\psi}_{i-1,n,m-1}\right)-g_2\left(\psi_{i-1,n,m-1}\right)\right|.\end{aligned} \tag{3c.2}$$

Assumption 3a suggests that for the first two terms on the right side of (3c.3)

$$E\left|g_j\left(\hat{\psi}_{i-1,n,m-1}\right)-g_j\left(\psi_{i-1,n,m-1}\right)\right|^2 \le E\left[\varphi_j^2\left(\psi_{i-1,n,m-1}\right)\left|\hat{\psi}_{i-1,n,m-1}-\psi_{i-1,n,m-1}\right|^2\right] \tag{3c.3}$$

where, in this case,

$$\begin{aligned} &E\left[\varphi_j^2\left(\psi_{i-1,n,m-1}\right)\left|\hat{\psi}_{i-1,n,m-1}-\psi_{i-1,n,m-1}\right|^2\right] \\ &\qquad \le E\left[\varphi_j^2\left(\psi_{i-1,n,m-1}\right)\left(\left|\hat{\psi}_{i-1,n,m-1}-\overline{\psi}_{i-1,n,m-1}\right|+\left|\overline{\psi}_{i-1,n,m-1}-\psi_{i-1,n,m-1}\right|\right)^2\right]. \end{aligned} \tag{3c.4}$$

Further expansion of the right side of (3c.4) leads to

$$\begin{aligned} &E\left[\varphi_j^2\left(\psi_{i-1,n,m-1}\right)\left(\left|\hat{\psi}_{i-1,n,m-1}-\overline{\psi}_{i-1,n,m-1}\right|+\left|\overline{\psi}_{i-1,n,m-1}-\psi_{i-1,n,m-1}\right|\right)^2\right] \\ &\qquad = E\left[\varphi_j^2\left(\psi_{i-1,n,m-1}\right)\left|\hat{\psi}_{i-1,n,m-1}-\overline{\psi}_{i-1,n,m-1}\right|^2\right]+E\left[\varphi_j^2\left(\psi_{i-1,n,m-1}\right)\left|\overline{\psi}_{i-1,n,m-1}-\psi_{i-1,n,m-1}\right|^2\right] \\ &\qquad + 2E\left[\varphi_j^2\left(\psi_{i-1,n,m-1}\right)\left|\hat{\psi}_{i-1,n,m-1}-\overline{\psi}_{i-1,n,m-1}\right|\left|\overline{\psi}_{i-1,n,m-1}-\psi_{i-1,n,m-1}\right|\right]. \end{aligned} \tag{3c.5}$$

Application of the Cauchy-Swarz Inequality then gives us

$$E\left[\varphi_j^2\left(\psi_{i-1,n,m-1}\right)\left|\hat{\psi}_{i-1,n,m-1}-\overline{\psi}_{i-1,n,m-1}\right|^2\right] \le E\left[\varphi_j^2\left(\psi_{i-1,n,m-1}\right)\right]E\left|\hat{\psi}_{i-1,n,m-1}-\overline{\psi}_{i-1,n,m-1}\right|^2, \tag{3c.6}$$

$$E\left[\varphi_j^2\left(\psi_{i-1,n,m-1}\right)\left|\overline{\psi}_{i-1,n,m-1}-\psi_{i-1,n,m-1}\right|^2\right] \le E\left[\varphi_j^2\left(\psi_{i-1,n,m-1}\right)\right]E\left|\overline{\psi}_{i-1,n,m-1}-\psi_{i-1,n,m-1}\right|^2, \tag{3c.7}$$

$$\begin{aligned} &E\left[\varphi_j^2\left(\psi_{i-1,n,m-1}\right)\left|\hat{\psi}_{i-1,n,m-1}-\overline{\psi}_{i-1,n,m-1}\right|\left|\overline{\psi}_{i-1,n,m-1}-\psi_{i-1,n,m-1}\right|\right] \\ &\qquad \le E\left[\varphi_j^2\left(\psi_{i-1,n,m-1}\right)\right]\left(E\left|\overline{\psi}_{i-1,n,m-1}-\psi_{i-1,n,m-1}\right|^2\right)^{1/2}\left(E\left|\hat{\psi}_{i-1,n,m-1}-\overline{\psi}_{i-1,n,m-1}\right|^2\right)^{1/2}. \end{aligned} \tag{3c.8}$$

The results of (3c.6), (3c.7) and (3c.8) suggest that we can bound the quantity in (3c.3) and, therefore, (3c.3) and (3c.3) as

$$\begin{aligned} &E\left|g_j\left(\hat{\psi}_{i-1,n,m-1}\right)-g_j\left(\psi_{i-1,n,m-1}\right)\right|^2 \le E\left[\varphi_j^2\left(\psi_{i-1,n,m-1}\right)\right] \\ &\cdot\left(E\left|\hat{\psi}_{i-1,n,m-1}-\overline{\psi}_{i-1,n,m-1}\right|^2+E\left|\overline{\psi}_{i-1,n,m-1}-\psi_{i-1,n,m-1}\right|^2+2\left(E\left|\overline{\psi}_{i-1,n,m-1}-\psi_{i-1,n,m-1}\right|^2 E\left|\hat{\psi}_{i-1,n,m-1}-\overline{\psi}_{i-1,n,m-1}\right|^2\right)^{1/2}\right), \end{aligned} \tag{3c.9}$$

$$\begin{aligned} &E\left|g_1\left(\hat{\psi}_{i-1,n,m-1}\right)-g_1\left(\psi_{i-1,n,m-1}\right)\right|^2 \le E\left[\varphi_1^2\left(\psi_{i-1,n,m-1}\right)\right] \\ &\left(E\left|\hat{\psi}_{i-1,n,m-1}-\overline{\psi}_{i-1,n,m-1}\right|^2+E\left|\overline{\psi}_{i-1,n,m-1}-\psi_{i-1,n,m-1}\right|^2+2\left(E\left|\overline{\psi}_{i-1,n,m-1}-\psi_{i-1,n,m-1}\right|^2 E\left|\hat{\psi}_{i-1,n,m-1}-\overline{\psi}_{i-1,n,m-1}\right|^2\right)^{1/2}\right), \end{aligned} \tag{3c.10}$$

and

$$\begin{aligned} &\gamma^2 E\left|g_2\left(\hat{\psi}_{i-1,n,m-1}\right)-g_2\left(\psi_{i-1,n,m-1}\right)\right|^2 \le \gamma^2 E\left[\varphi_2^2\left(\psi_{i-1,n,m-1}\right)\right] \\ &\left(E\left|\hat{\psi}_{i-1,n,m-1}-\overline{\psi}_{i-1,n,m-1}\right|^2+E\left|\overline{\psi}_{i-1,n,m-1}-\psi_{i-1,n,m-1}\right|^2+2\left(E\left|\overline{\psi}_{i-1,n,m-1}-\psi_{i-1,n,m-1}\right|^2 E\left|\hat{\psi}_{i-1,n,m-1}-\overline{\psi}_{i-1,n,m-1}\right|^2\right)^{1/2}\right), \end{aligned} \tag{3c.11}$$

respectively.

Similarly, Assumption 3a and the Triangle Inequality suggest that we have for the final term on the right side of (3c.3)

$$E\left|g_1\left(\hat{\psi}_{i-1,n,m-1}\right)-g_1\left(\psi_{i-1,n,m-1}\right)\right|\left|g_2\left(\hat{\psi}_{i-1,n,m-1}\right)-g_2\left(\psi_{i-1,n,m-1}\right)\right| \\ \leq\left(E\left|g_1\left(\hat{\psi}_{i-1,n,m-1}\right)-g_1\left(\psi_{i-1,n,m-1}\right)\right|^2\right)^{1/2}\left(E\left|g_2\left(\hat{\psi}_{i-1,n,m-1}\right)-g_2\left(\psi_{i-1,n,m-1}\right)\right|^2\right)^{1/2}. \tag{3c.12}$$

Therefore, we can conclude, using the results in (3c.9), that

$$E\left|g_1\left(\hat{\psi}_{i-1,n,m-1}\right)-g_1\left(\psi_{i-1,n,m-1}\right)\right|\left|g_2\left(\hat{\psi}_{i-1,n,m-1}\right)-g_2\left(\psi_{i-1,n,m-1}\right)\right|\leq\left(E\left[\varphi_1^2\left(\psi_{i-1,n,m-1}\right)\right]E\left[\varphi_2^2\left(\psi_{i-1,n,m-1}\right)\right]\right)^{1/2} \\ \left(E\left|\hat{\psi}_{i-1,n,m-1}-\overline{\psi}_{i-1,n,m-1}\right|^2+E\left|\overline{\psi}_{i-1,n,m-1}-\psi_{i-1,n,m-1}\right|^2+2\left(E\left|\overline{\psi}_{i-1,n,m-1}-\psi_{i-1,n,m-1}\right|^2E\left|\hat{\psi}_{i-1,n,m-1}-\overline{\psi}_{i-1,n,m-1}\right|^2\right)^{1/2}\right) \tag{3c.13}$$

The results in (3c.10), (3c.11) and (3c.14) suggest that we bound the quantity of interest in (3c.1) as

$$\begin{aligned}
&E\left|\overline{\psi}_{i,n,m}-\psi_{i,n,m}\right|^2\leq E\left[\varphi_1^2\left(\psi_{i-1,n,m-1}\right)\right] \\
&\left(E\left|\hat{\psi}_{i-1,n,m-1}-\overline{\psi}_{i-1,n,m-1}\right|^2+E\left|\overline{\psi}_{i-1,n,m-1}-\psi_{i-1,n,m-1}\right|^2+2\left(E\left|\overline{\psi}_{i-1,n,m-1}-\psi_{i-1,n,m-1}\right|^2E\left|\hat{\psi}_{i-1,n,m-1}-\overline{\psi}_{i-1,n,m-1}\right|^2\right)^{1/2}\right) \\
&\qquad+\gamma^2E\left[\varphi_2^2\left(\psi_{i-1,n,m-1}\right)\right] \\
&\left(E\left|\hat{\psi}_{i-1,n,m-1}-\overline{\psi}_{i-1,n,m-1}\right|^2+E\left|\overline{\psi}_{i-1,n,m-1}-\psi_{i-1,n,m-1}\right|^2+2\left(E\left|\overline{\psi}_{i-1,n,m-1}-\psi_{i-1,n,m-1}\right|^2E\left|\hat{\psi}_{i-1,n,m-1}-\overline{\psi}_{i-1,n,m-1}\right|^2\right)^{1/2}\right) \\
&\qquad+2|\gamma|\left(E\left[\varphi_1^2\left(\psi_{i-1,n,m-1}\right)\right]E\left[\varphi_2^2\left(\psi_{i-1,n,m-1}\right)\right]\right)^{1/2} \\
&\left(E\left|\hat{\psi}_{i-1,n,m-1}-\overline{\psi}_{i-1,n,m-1}\right|^2+E\left|\overline{\psi}_{i-1,n,m-1}-\psi_{i-1,n,m-1}\right|^2+2\left(E\left|\overline{\psi}_{i-1,n,m-1}-\psi_{i-1,n,m-1}\right|^2E\left|\hat{\psi}_{i-1,n,m-1}-\overline{\psi}_{i-1,n,m-1}\right|^2\right)^{1/2}\right),
\end{aligned} \tag{3c.14}$$

which can be conveniently simplified to

$$E\left|\overline{\psi}_{i,n,m}-\psi_{i,n,m}\right|^2\leq D_{i-1,n,m-1} \\ \left(E\left|\hat{\psi}_{i-1,n,m-1}-\overline{\psi}_{i-1,n,m-1}\right|^2+E\left|\overline{\psi}_{i-1,n,m-1}-\psi_{i-1,n,m-1}\right|^2+2\left(E\left|\overline{\psi}_{i-1,n,m-1}-\psi_{i-1,n,m-1}\right|^2E\left|\hat{\psi}_{i-1,n,m-1}-\overline{\psi}_{i-1,n,m-1}\right|^2\right)^{1/2}\right) \tag{3c.15}$$

where $D_{i-1,n,m-1}=E\left[\varphi_1^2\left(\psi_{i-1,n,m-1}\right)\right]+E\left[\varphi_2^2\left(\psi_{i-1,n,m-1}\right)\right]+2\left(E\left[\varphi_1^2\left(\psi_{i-1,n,m-1}\right)\right]E\left[\varphi_2^2\left(\psi_{i-1,n,m-1}\right)\right]\right)^{1/2}$.

Equivalently, by taking square root on both sides of (3c.15), and referring to the norms of the vectors, the inequality could be written as

$$\left\|\bar{\psi}_{i,n,m}-\psi_{i,n,m}\right\|_2 \le d_{i-1,n,m-1} \left(\left\|\hat{\psi}_{i-1,n,m-1}-\bar{\psi}_{i-1,n,m-1}\right\|_2+\left\|\bar{\psi}_{i-1,n,m-1}-\psi_{i-1,n,m-1}\right\|_2+2\left(\left\|\hat{\psi}_{i-1,n,m-1}-\bar{\psi}_{i-1,n,m-1}\right\|_2\left\|\bar{\psi}_{i-1,n,m-1}-\psi_{i-1,n,m-1}\right\|_2\right)^{1/2}\right) \tag{3c.16}$$

where

$$\begin{aligned}\left(D_{i-1,n,m-1}\right)^{1/2} \ge d_{i-1,n,m-1} &= \left\|\varphi_1\left(\psi_{i-1,n,m-1}\right)\right\|_2+\left\|\varphi_2\left(\psi_{i-1,n,m-1}\right)\right\|_2+2\left(\left\|\varphi_1\left(\psi_{i-1,n,m-1}\right)\right\|_2\left\|\varphi_2\left(\psi_{i-1,n,m-1}\right)\right\|_2\right)^{1/2}\\ &= \left(\left(\left\|\varphi_1\left(\psi_{i-1,n,m-1}\right)\right\|_2\right)^{1/2}+\left(\left\|\varphi_2\left(\psi_{i-1,n,m-1}\right)\right\|_2\right)^{1/2}\right)^2.\end{aligned} \tag{3c.17}$$

Furthermore, performing backward iteration on (3c.16) will then lead to

$$\begin{aligned}&\left\|\bar{\psi}_{i,n,m}-\psi_{i,n,m}\right\|_2\\ &\le d_{i-1,n,m-1}\left(\left\|\hat{\psi}_{i-1,n,m-1}-\bar{\psi}_{i-1,n,m-1}\right\|_2+2\left(\left\|\hat{\psi}_{i-1,n,m-1}-\bar{\psi}_{i-1,n,m-1}\right\|_2\left\|\bar{\psi}_{i-1,n,m-1}-\psi_{i-1,n,m-1}\right\|_2\right)^{1/2}\right)\\ &+d_{i-1,n,m-1}d_{i-2,n,m-2}\left(\left\|\hat{\psi}_{i-2,n,m-2}-\bar{\psi}_{i-2,n,m-2}\right\|_2+2\left(\left\|\hat{\psi}_{i-2,n,m-2}-\bar{\psi}_{i-2,n,m-2}\right\|_2\left\|\bar{\psi}_{i-2,n,m-2}-\psi_{i-2,n,m-2}\right\|_2\right)^{1/2}\right)\\ &+\ldots+d_{i-1,n,m-1}d_{i-2,n,m-2}\cdots d_{i-(m-1),n,1}\left(\left\|\hat{\psi}_{i-(m-1),n,1}-\bar{\psi}_{i-(m-1),n,1}\right\|_2+2\left(\left\|\hat{\psi}_{i-(m-1),n,1}-\bar{\psi}_{i-(m-1),n,1}\right\|_2\left\|\bar{\psi}_{i-(m-1),n,1}-\psi_{i-(m-1),n,1}\right\|_2\right)^{1/2}\right)\\ &+d_{i-1,n,m-1}\cdots d_{i-(m-1),n,1}d_{i-m,n,0}\left(\left\|\hat{\psi}_{i-m,n,0}-\bar{\psi}_{i-m,n,0}\right\|_2+\left\|\bar{\psi}_{i-m,n,0}-\psi_{i-m,n,0}\right\|_2+2\left(\left\|\hat{\psi}_{i-m,n,0}-\bar{\psi}_{i-m,n,0}\right\|_2\left\|\bar{\psi}_{i-m,n,0}-\psi_{i-m,n,0}\right\|_2\right)^{1/2}\right)\end{aligned}$$

where $d_{i-s,n,m-s}=\left(\left(\left\|\varphi_1\left(\psi_{i-s,n,m-s}\right)\right\|_2\right)^{1/2}+\left(\left\|\varphi_2\left(\psi_{i-s,n,m-s}\right)\right\|_2\right)^{1/2}\right)^2$ for $s=1,\cdots,m$. (3c.18)

Now, observe that because $\left\{\psi_{i,n,0};1\le i\le T\right\}$ are the starting values, therefore, it is reasonable to assume

$$\hat{\psi}_{i-m,n,0}=\bar{\psi}_{i-m,n,0}=\psi_{i-m,n,0}. \tag{3c.19}$$

Therefore, using the definitions in (3.14), the results in (3c.18) can be simplified to

$$\left\|\bar{\psi}_{i,n,m}-\psi_{i,n,m}\right\|_2 \le \left(d_{i-1,n,m-1}+d_{i-1,n,m-1}d_{i-2,n,m-2}+\ldots+d_{i-1,n,m-1}d_{i-2,n,m-2}\cdots d_{i-(m-1),n,1}\right)\Delta_{1\kappa}^{(L_2)}\left(1+2\frac{\left(\Delta_{2\kappa}^{(L_2)}\right)^{1/2}}{\left(\Delta_{1\kappa}^{(L_2)}\right)^{1/2}}\right). \tag{3c.20}$$

Finally, by using the second part of Assumption 3a, we are able to summarise (3c.30) as

$$\left\|\bar{\psi}_{i,n,m}-\psi_{i,n,m}\right\|_2 \le \sum_{j=1}^{m-1} G^j \Delta_{1\kappa}^{(L_2)}\left\{1+2\left(\frac{\Delta_{2\kappa}^{(L_2)}}{\Delta_{1\kappa}^{(L_2)}}\right)^{1/2}\right\} \tag{3c.21}$$

for some $0<G<1$.

Proof of Theorem 3d

The decomposition in (3.9), and the results of Theorems 3b and 3c collectively suggest that we have

$$\left\|\hat{\psi}_{i,n,m}-\psi_i\right\|_2 \le \left\|\hat{\psi}_{i,n,m}-\bar{\psi}_{i,n,m}\right\|_2+\sum_{j=1}^{m-1}G^j\Delta_{1\kappa}^{(L_2)}\left\{1+2\left(\frac{\Delta_{2\kappa}^{(L_2)}}{\Delta_{1\kappa}^{(L_2)}}\right)^{1/2}\right\}+G^m\left(V_1+V_2\right) \tag{3d.1}$$

for some $0<G<1$. It is obvious from the proof of Theorem 3a that statistically $\Delta_{1\kappa}^{(L_2)}$ in (3d.1) can be extended to

$$\Delta_{1m}^{(L_2)} =: \sup_{m\ge 1}\left\|\hat{\psi}_{i,m}-\bar{\psi}_{i,m}\right\|_2, \tag{3d.2}$$

where $\hat{\psi}_{i,m}=\left(\hat{\psi}_{i,1,m},\dots,\hat{\psi}_{i,N,m}\right)^{\tau}$ and $\bar{\psi}_{i,m}=\left(\bar{\psi}_{i,1,m},\dots,\bar{\psi}_{i,N,m}\right)^{\tau}$. The above result suggests that (3d.1) can be conveniently rewritten as

$$\begin{aligned}
\left\|\hat{\psi}_{i,n,m}-\psi_i\right\|_2 &\le \Delta_{1m}^{(L_2)}+\sum_{j=1}^{m-1}G^j\Delta_{1m}^{(L_2)}\left\{1+2\left(\frac{\Delta_{2\kappa}^{(L_2)}}{\Delta_{1\kappa}^{(L_2)}}\right)^{1/2}\right\}+G^m\left(V_1+V_2\right)\\
&=\sum_{j=0}^{m-1}G^j\Delta_{1m}^{(L_2)}+2\sum_{j=0}^{m-1}G^j\Delta_{1m}^{(L_2)}\left(\frac{\Delta_{2\kappa}^{(L_2)}}{\Delta_{1\kappa}^{(L_2)}}\right)^{1/2}-2\Delta_{1m}^{(L_2)}\left(\frac{\Delta_{2\kappa}^{(L_2)}}{\Delta_{1\kappa}^{(L_2)}}\right)^{1/2}+G^m\left(V_1+V_2\right)\\
&=\frac{\Delta_{1m}^{(L_2)}}{(1-G)}\left\{1+2\left(\frac{\Delta_{2\kappa}^{(L_2)}}{\Delta_{1\kappa}^{(L_2)}}\right)^{1/2}\right\}-2\Delta_{1m}^{(L_2)}\left(\frac{\Delta_{2\kappa}^{(L_2)}}{\Delta_{1\kappa}^{(L_2)}}\right)^{1/2}+G^m\left(V_1+V_2\right).
\end{aligned} \tag{3d.3}$$

In addition, a few steps of manipulation would lead to a much simpler form of (3d.3) such that

$$\begin{aligned}
\left\|\hat{\psi}_{i,n,m}-\psi_i\right\|_2 &\le \frac{\Delta_{1m}^{(L_2)}}{(1-G)}\left\{1+2G\left(\frac{\Delta_{2\kappa}^{(L_2)}}{\Delta_{1\kappa}^{(L_2)}}\right)^{1/2}\right\}+G^m\left(V_1+V_2\right)\\
&\equiv A_1+A_2
\end{aligned} \tag{3d.4}$$

for some $0<G<1$. While it is clear that A_2 converges to zero as $m\to\infty$, using the results of Theorem 3a we can also conclude that

$$A_1=O\left(\Delta_{1m}^{(L_2)}\right) \tag{3d.5}$$

and, therefore, that

$$\left\|\hat{\psi}_{i,n,m}-\psi_i\right\|_2^2=O\left(\Delta_{1m}^{(L_2)}\right)^2=O\left(\frac{c_1}{(T-1)h}+c_2h^4+o_P\left\{\frac{1}{(T-1)h}+h^4\right\}\right) \textit{ uniformly over } h\in H_T, \tag{3d.6}$$

which complete the proof of Theorem 3d.

References

- Allen, D., Chan, C., McAleer, M., and Peiris, S. (2006). Finite Sample Properties of the QMLE for the Log-ACD Model: Application to Australian Stocks. Working Paper, FIMARC, Edith Cowan University.
- An, H. A., and Huang, F. C. (1996). The Geometrical Ergodicity of Nonlinear Autoregressive Models. *Statistica Sinica*, 6, 943-956.
- Bauwens, L., and Giot, P. (2000). The Logarithmic ACD Model: An Application to the Bid-Ask Quote Process of Three NYSE Stocks. *Annales d'Economie et de Statistique*, 60, 117-149.
- Bauwens, L., Grammig, J., Veredas, D., and Giot, P. (2004). A Comparison of Financial Duration Models Via Density Forecast. *International Journal of Forecasting*, 20, 589-604.
- Bollerslev, T. (1986). Generalised Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31, 307-327.
- Buhlmann, P. and McNeil, A. J. (2002). An Algorithm for Nonparametric GARCH Modelling. *Computational Statistics and Data Analysis*, 40, 665-683.
- Dufour, A., and Engle, R. F. (2000). Time and the Impact of a Trade. *Journal of Finance*, 55, 6, 2467-2498.
- Easley, D., and O'Hara, M. (1992). Time and the Process of Security Price Adjustment. *Journal of Finance*, 47, 905-927.
- Engle, R. F., Granger, C. W. J., Rice, J. A., and Weiss, A. (1986). Semiparametric Estimates of the Relation between Weather and Electricity Sale. *Journal of the American Statistical Association*, 81, 310-320.
- Engle, R. F., and Russell, J. R. (1998). Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. *Econometrica*, 66, 1127-1162.
- Fernandes, M., and Grammig, J. (2006). A Family of Autoregressive Conditional Duration Models. *Journal of Econometrics*, 130, 1-23.
- Fernandes, M., Medeiros, M. C., and Veiga, A. (2006). A (Semi-) Parametric Functional Coefficient Autoregressive Conditional Duration Model. Working Papers, Department of Economics, PUC-Rio.

- Gao, J. (2006). *Nonlinear Time Series: Semiparametric Methods*. London: Chapman and Hall/CRC.
- Gao, J., and Yee, T. (2000). Adaptive Estimation in Partially Linear Autoregression Models. *Canadian Journal of Statistics*, 28, 571-586.
- GyÖrfi, L., Härdle, W., Sarda, P., and Vieu, P. (1989). *Nonparametric Curve Estimation for Time Series, Lecture Notes in Statistics*. Springer: New York.
- Härdle , W., and Vieu, P. (1992). Kernel Regression Smoothing of Time Series. *Journal of Time Series Analysis*, 13, 209-232.
- Hardle, W., Liang, H., and Gao J. (2000). *Partially Linear Models*. New York: Springer Series: Contributions to Statistics. Physica-Verlag.
- Hasbrouck, J. (1991). Measuring the Information Content of Stock Trades, *Journal of Finance*, 66, 1, 179-207.
- Heckman, N. E. (1986). Spline Smoothing of Time Series. *Journal Time Series Anal.*, 13, 209-232.
- Lee, S. and Hansen, B. (1994). Asymptotic Theory for the GARCH(1,1) Quasi-Maximum Likelihood Estimator, *Econometric Theory*, 10, 29-52.
- Meitz, M., and Teräsvirta , T. (2006). Evaluating models of Autoregressive Conditional Duration. Journal of Business and Economic Statistics, 24, 104-124.
- Nychka, D., Elliner, S., Gallange, A. and McCaffrey, D. (1992). Finding Chaos in Noisy Systems. *Journal of Royal Statistics Society*. B, 54, 399-426.
- Tsay, R. S. (2005). *Analysis of Financial Time Series*. New York: Wiley.
- Vuorenmaa, T. A. (2006). A q-Weibull Autoregressive Conditional Duration Model. Proceedings from TSEFR'06: International Conference on Time Series Econometrics, Finance and Risk. Perth, Western Australia.
- Zhang, M. Y., Russell, J. R., and Tsay, R. S. (2001). A Nonlinear Autoregressive Conditional Duration Model with Application to Financial Transaction Data. *Journal of Econometrics*, 104, 179-207.